

**Design, syntheses and performance evaluation
of novel generations of lanthanide chelating
tags and an NMR study of formylglycine-
generating enzyme**

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Daniel Joss

Basel, 2021

Originaldokument gespeichert auf dem Dokumentenserver der Universität Basel

edoc.unibas.ch

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät:

auf Antrag von

Prof. Dr. Daniel Häussinger

Prof. Dr. Florian P. Seebeck

Prof. Dr. Marcellus Ubbink

Basel, den 22.06.2021

Prof. Dr. Marcel Mayor, Dekan

*“One, remember to look up at the stars and not down at your feet.
Two, never give up work. Work gives you meaning and purpose and life is empty without it.
Three, if you are lucky enough to find love, remember it is there and don't throw it away.”*

Stephen Hawking

Abstract

In order to provide cutting-edge tools for probing structure and dynamics of proteins by the use of lanthanide chelating tags (LCTs), new generations of LCTs with novel structural features and enhanced properties were developed. First, the presence of multiple conformational species and the averaging of paramagnetic effects by motional flexibility of the LCT on the surface of the protein was addressed. Subsequently, the development of an activator moiety that leads to a short, reduction-stable linkage formed upon tagging of the protein yielded an LCT with unprecedented induced anisotropy parameters and favorable ligation kinetics. Furthermore, the design and synthesis of a novel LCT that is suitable for tagging of synthetic RNA strands containing a modified nucleobase extension was achieved. Thereby, the repertoire of paramagnetic nuclear magnetic resonance (NMR) spectroscopy using LCTs was extended from the protein to the ribonucleic acid (RNA) world. Additionally, a comprehensive study about the formylglycine-generating enzyme (FGE) with its unusual fold and catalytic properties and the investigation of four novel antigens in cancer research was undertaken. An abstract of the five parts of this thesis is given in the following.

In the first part of this thesis, improved generations of LCTs are included. In order to push the limits in paramagnetic NMR spectroscopy using LCTs, one existing and three completely novel 1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid (DOTA)-derived LCTs were extensively studied. The exploration of the field started with the synthesis and performance evaluation of the eight-fold methyl-substituted Ln-DOTA-M8-(4*R*4*S*)-SSPy (DOTA-M8-(4*R*4*S*)-SSPy: (2*R*,2'*R*,2''*R*)-2,2',2''-((2*S*,5*S*,8*S*,11*S*)-2,5,8,11-tetramethyl-10-((*R*)-1-oxo-1-(2-(pyridin-2-yl)disulfanyl)ethylamino)propan-2-yl)-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)tripropionate) that is locked in only one conformational species. Subsequently, an LCT with four isopropyl substituents in the basic ring scaffold, i.e. Ln-P4M4-DOTA, was designed and synthesized (P4M4-DOTA: (2*R*,2'*R*,2''*R*)-2,2',2''-((2*S*,5*S*,8*S*,11*S*)-2,5,8,11-tetraisopropyl-10-((*R*)-1-oxo-1-(2-(pyridin-2-yl)disulfanyl)-ethylamino)propan-2-yl)-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)tripropionate). Both the Tm- and Dy- complex of the novel LCT displayed strongly enhanced pseudocontact shifts (PCSs) on both ubiquitin and hCA II constructs when compared to Ln-DOTA-M8-(4*R*4*S*)-SSPy. The isopropyl-substitution pattern was then combined with a reduction-stable linker suitable for *in cell* applications (P4T-DOTA: (2*R*,2'*R*,2''*R*)-2,2',2''-((2*S*,5*S*,8*S*,11*S*)-2,5,8,11-tetraiso-propyl-10-((2-(methylsulphonyl)thiazolo[5,4-*b*]pyridine-5-yl)methyl)-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)tripropionate). Due to the combination of the isopropyl substituents on the basic ring scaffold and the short linker moiety, the obtained PCSs and residual dipolar couplings (RDCs) were further increased when compared to Ln-DOTA-M8-(4*R*4*S*)-SSPy and Ln-P4M4-DOTA. The different stages of

development then culminated in the rational design of Ln-M7-Nitro (M7-Nitro: Ln-(2*R*,2'*R*,2''*R*)-2,2',2''-((2*S*,5*S*,8*S*,11*S*)-2,5,8,11-tetramethyl-10-((6-(methylsulphonyl)-5-nitropyridin-2-yl)-methyl)-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)tripropanoate), a novel LCT exhibiting unprecedented anisotropy parameters on ubiquitin S57C, a reduction-stable linker, and strongly improved ligation kinetics, i.e. tagging of the cysteine residue of the investigated protein constructs to an extent greater than 95% over 30 minutes under close to physiological conditions (highly diluted (90 μ M), aqueous buffer, pH 7.0). The favorable properties of the Ln-M7-Nitro LCT are mainly based on its novel, highly reactive and selective 5-nitro-substituted 6-(methylsulphonyl)-2-methylpyridine activator moiety.

The second part of the thesis comprises a review about the design and applications of LCTs for the investigation of proteins, nucleic acids, carbohydrate and complexes that was published in *Progress in Nuclear Magnetic Resonance Spectroscopy*. Subsequently, a review about the application of LCTs for the localization of ligands within biomacromolecules is included that will contribute as book chapter to *Comprehensive Coordination Chemistry III*.

The third part of this thesis deals with the transfer of the rich LCT toolbox from proteins to RNA. RNA molecules become increasingly important for future drug design and there is currently no LCT available that enables a successful covalent conjugation to RNA strands and subsequent exploitation of the paramagnetic effects induced by the lanthanide ions for structure elucidation. Therefore, an LCT comprising a novel activator moiety was designed and benchmarked on a 36-nt RNA strand that incorporates a modified nucleobase. The detected anisotropic effects with the newly designed LCT correspond within a range of 10% to those obtained with Ln-DOTA-M8-(4*R*4*S*)-SSPy on an ubiquitin S57C construct and it can therefore be concluded that the PCS methodology was successfully transferred from the protein to the RNA world and will contribute to future studies of structure, dynamics and ligand-binding of RNA molecules in solution.

In the fourth part of the thesis, a solution NMR study of structure, dynamics and mechanism of the FGE protein is included. The study yields residue-specific insights into the solution state structure and dynamics of a uniformly ^2H , ^{13}C , ^{15}N labeled FGE construct and a uniformly ^{15}N labeled FGE construct by use of amide proton exchange, T_1 and T_2 relaxation and heteronuclear Overhauser effect NMR experiments. Furthermore, the catalytic site of the enzyme was investigated by following chemical shift perturbations (CSPs) in ^1H , ^{15}N -HSQC experiments recorded with samples of the apoenzyme, Cu(I)- and Ag(I)-containing enzyme as well as with several mutations and with and without substrate loading. Subsequently, the CSPs in ^1H , ^{15}N -HSQC experiments recorded on a disease-related mutant of FGE were analyzed in

order to provide a molecular basis for the occurrence of specific phenotypes in patients with multiple sulfatase deficiency (MSD). The solution NMR study extends the current knowledge about this highly special and exciting copper-containing enzyme.

The fifth and last part of this thesis comprises a study about a novel class of antigens in cancer research. Four novel antigens were synthesized and structurally characterized by NMR and HRMS. The biological experiments showed that the antigens stimulate T cells and can therefore be exploited for activation of specific T lymphocytes in order to recognize cancer cells. This novel class of antigens offers thereby an enormous therapeutic potential for the development of a potential vaccine against different types of cancer.

Publications

12. "Conference Report - Christmas Symposium Basel."
D. Joss, A. Baiyoumy, V. Chimisso, C. E. Meyer, D. A. Miarzlou, G. M. Murawska, N. Niggli, B. Pfund, K. Reznikova, P. Rieder, *Chimia*, **2021**, 75, 112.
11. "A novel, rationally designed lanthanoid chelating tag delivers large paramagnetic structural restraints for biomolecular NMR."
D. Joss, F. Winter, D. Häussinger, *Chemical Communications*, **2020**, 56, 12861-12864.
10. "Application of paramagnetic lanthanoid chelating tags in NMR spectroscopy and their use for the localization of ligands within biomacromolecules."
D. Joss, R. Vogel, K. Zimmermann, D. Häussinger, *Reference Module in Chemistry, Molecular Sciences and Chemical Engineering*, **2020**, doi:10.1016/B978-0-12-409547-2.14848-6.
9. "Conference Report - Christmas Symposium Basel."
D. Joss, J. Vallapurackal, A. Baiyoumy, V. Chimisso, F. Christoffel, P. Herr, C. Meyer, D. Miarzlou, G. Murawska, N. Niggli, K. Reznikova, I. Urosev, P. van Gerwen, *Chimia*, **2020**, 74, 63.
8. "Design and application of lanthanide chelating tags for pseudocontact shift nuclear magnetic resonance spectroscopy on biomacromolecules."
D. Joss, D. Häussinger, *Progress in Nuclear Magnetic Resonance Spectroscopy*, **2019**, 114-115, 284-312.
7. "P4T-DOTA - A lanthanide chelating tag combining a highly sterically overcrowded backbone with a reductively stable linker."
D. Joss, D. Häussinger, *Chemical Communications*, **2019**, 55, 10543-10546.
6. "Structure of formylglycine-generating enzyme in complex with copper and a substrate reveals an acidic pocket for binding and activation of molecular oxygen."
D. A. Miarzlou, F. Leisinger, D. Joss, D. Häussinger, F. P. Seebeck, *Chemical Science*, **2019**, 10, 7049-7058.

5. "A sterically overcrowded, isopropyl-substituted lanthanide chelating tag for protein PCS NMR spectroscopy: Synthesis of its macrocyclic scaffold and benchmarking on ubiquitin S57C and hCA II S166C."
D. Joss, M.-S. Bertrams, D. Häussinger, *Chemistry - A European Journal*, **2019**, 25, 11910-11917.
4. "Localization of ligands within human carbonic anhydrase II using ¹⁹F pseudocontact shift analysis."
K. Zimmermann, D. Joss, T. Müntener, E. S. Nogueira, M. Schäfer, L. Knörr, F. W. Monnard, D. Häussinger, *Chemical Science*, **2019**, 10, 5064-5072.
3. "Synthesis of chiral nine and twelve-membered cyclic polyamines from natural building blocks."
T. Müntener, F. Thommen, D. Joss, J. Kottelat, A. Prescimone, D. Häussinger, *Chemical Communications*, **2019**, 32, 4715-4718.
2. "Conference Report - Christmas Symposium Basel."
D. Joss, S. Rigo, F. Christoffel, P. Herr, G. Murawska, B. Sauter, A. Stampfli, I. Urosev, J. Vallapurackal, *Chimia*, **2019**, 73, 98.
1. "Conformationally locked lanthanide chelating tags for convenient pseudocontact shift protein nuclear magnetic resonance spectroscopy."
D. Joss, R. M. Walliser, K. Zimmermann, D. Häussinger, *Journal of Biomolecular NMR*, **2018**, 72, 29-38.

Acknowledgements

I am very grateful to my supervisor Prof. Dr. Daniel Häussinger for all the support and guidance during my PhD. Besides the deep insights that you offered me concerning the theoretical and practical aspects of NMR spectroscopy, I very much appreciated all the confidence and trust in me and the thereby associated freedom in my research projects. Besides the excellent mentoring in all chemical and biological aspects of the projects, I also learned a lot about strategic and personal aspects in life and what it means to guide people by example with both deep knowledge and passion. The trips to the Swiss alps and conferences as well as the cohesiveness of the research group also in stressful times as during a pandemic make the PhD period one of the best in my entire life. Thank you for all your past and future support!

Prof. Dr. Florian P. Seebeck impressed me with his thorough knowledge about protein science as well as about all aspects of chemistry in general and I am very thankful for the opportunity to have grown as a PhD student with you as second supervisor/co-examiner, Fakultätsverantwortlicher and collaborator in the FGE projects. You always had an open ear for me both as NMR spectroscopist but also as person. To discuss cutting-edge research including NMR relaxation and structural properties of protein loops in “Bärndütsch” is just great!

Prof. Dr. Marcellus Ubbink is thanked for his careful evaluation of my thesis as external expert.

I would like to thank Prof. Dr. Dennis Gillingham for his time and effort in being the chairman of my PhD defense.

Dzmitry Miarzlou is thanked for the very pleasant collaboration in the FGE research projects. This projects would have never been that successful without such an excellent coworker!

Prof. Dr. Christoph Kreutz is thanked for the very fruitful collaboration in the RNA project. The time spent in Innsbruck with you, your group and especially Kevin Erharter was very pleasant both in terms of scientific as well as personal exchange and provided many new and very important results.

I would like to thank Qinmei Yang, Alessandro Vacchini, Andrew Chancellor and Prof. Dr. Gennaro De Libero for the highly interesting, efficient and fruitful collaboration project concerning T cell activation by nucleobase adduct-containing metabolites!

Florine Winter is gratefully acknowledged for her endless motivation and efforts in the development of the Ln-M7-Nitro LCT during her master thesis. To teach and guide a master student like you was just a pleasure!

I would like to thank Maria-Sophie Bertrams, Roman Misteli and Sandro Fischer for performance of their “Schlussversuch” under my supervision!

Thomas Müntener, Raphael Vogel and Pascal Rieder are very much thanked for both the scientifically and personally highly enriching atmosphere in the lab during my PhD. It was a privilege to interact with you not only in many scientifically very helpful and well-founded discussions, but also just in funny and non-sense discussions during the coffee breaks and in the evenings.

I would also like to thank all former members of the Häussinger group, i.e. Florine Winter, Fabienne Thommen, Dr. Thomas Müntener, Florian Lüttin, Dr. Roché M. Walliser, Dr. Kaspar Zimmermann and Dr. Heiko Gsellinger for their encouragement and the pleasant alumni/research group meetings.

All current and past members of the PhD chemistry community (PCC) are gratefully acknowledged. The organization of both the social and scientific activities was great fun and I will never forget our ski weekends in the Swiss alps.

Special thanks go to my flatmate and research group “neighbor” David Vogel for his friendship and all the interesting and fun discussions about life and science!

Dr. Thomas Müntener is gratefully acknowledged for his guidance as senior PhD student in the first year of my PhD and for proof-reading my PhD thesis during his postdoc.

Special thanks go to the Mayor group for the open atmosphere and inclusion in your activities. Furthermore, I would like to thank Adriano D’Addio for the support with truly excellent coffee!

I would like to thank all the technical and administrative staff of the chemistry department for their support. The foundations for successful research would not be given without your important work!

Dr. Michael Pfeffer is acknowledged for measuring HRMS samples and the excellent collaboration in the antigen project.

The department of chemistry is gratefully acknowledged for financial support. Without the excellent facilities and financial support, the research in this thesis could not have been done.

All members of the team of the university's high performance computing facility sciCORE are acknowledged for providing the technical possibilities for DFT calculations.

I would like to express my deepest gratitude to my parents, Anne and Bernhard, and to my sister Helen! You were just an awesome support in the last 28 years and I would not be where I am now without your endless support and thoughtful advice. I would also like to thank Johanna for all the good times we shared during the PhD! Besides the ski tours in the beautiful Swiss alps and the extensive hiking tours, I very much enjoyed all your support and your refreshing views on science and life.

Last but not least, I would like to thank all the members of the chemistry department and all people, that I might have forgotten to mention explicitly. You contributed so much to my PhD and without you the chemistry department of the University of Basel would not be the awesome institute that it is!

Acronyms

3MDPA	3-mercapto-2,6-pyridinedicarboxylic acid
4-HNE	4-hydroxy-nonenal
4MDPA	4-mercapto-2,6-pyridinedicarboxylic acid
4MMDPA	4-mercaptomethyl-dipicolinic acid
4MTDA	4'-mercapto-2,2': 6',2''-terpyridine-6,6''-dicarboxylic acid
4PS-DPA	(4-(phenylsulfonyl)-pyridine-2,6-dicarboxylic acid
4PS-PyMTA	4-phenylsulfonyl-(pyridin-2,6-diyl)bismethylenenitrilotetrakis(acetic acid)
4VPyMTA	4-vinyl(pyridine-2,6-diyl)bismethylenenitrilo tetrakis(acetic acid)
5-OP-RU	5-(2-oxopropylideneamino)-6-d-ribitylaminouracil
7TM	seven-helix transmembrane
ADA	adenosine deaminase
ADP	adenosine diphosphate
ADSSL1	adenylosuccinate synthase 1
Ag	antigen
AIBN	azobisisobutyronitrile
ALDH16A1	aldehyde dehydrogenase 16 family member A1
APC	anaphase-promoting complex
ArgN	N-terminal domain of the arginine repressor from <i>E. coli</i>
ATR	attenuated total reflectance
AzF	<i>p</i> -azido- <i>L</i> -phenylalanine
B2M	beta-2-microglobulin
BBG	<i>S-p</i> -bromobenzylglutathione
BRM	Brahma (<i>Drosophila</i> protein)
BSO	buthionine sulfoximine
BTT	5-(benzylthio)-1 <i>H</i> -tetrazole
BTTAA	2-(4-((bis((1-(<i>tert</i> -butyl)-1 <i>H</i> -1,2,3-triazol-4-yl)methyl)amino)methyl)-1 <i>H</i> -1,2,3-triazol-1-yl)acetic acid
CAO	copper amine oxidase
CARA	computer-aided resonance assignment
Cas9	CRISPR-associated endonuclease
CAT	catalase
Cbz	carboxybenzyl
CD11a	integrin α -L
CD	cluster of differentiation

CEST	chemical exchange saturation transfer
cGMP	cyclic guanosine monophosphate
CLaNP	caged lanthanide NMR probe
COSY	correlation spectroscopy
CPCM	conductor-like polarizable continuum model
CRISPR	clustered regularly interspaced short palindromic repeats
Cryo-EM	cryogenic electron microscopy
CS	contact shift
CSP	chemical shift perturbation
CTIP	2-cyanoethyl- <i>N,N</i> -diisopropylchlorphosphoramidite
CueR	copper efflux regulator
Cyt c	cytochrome c
dAdo	deoxyadenosine
DBH	dopamine- β -hydroxylase
DCM	dichloromethane
DEER	double electron-electron resonance
DENpro	dengue virus type 2 NS2B/NS3 protease
DFT	density functional theory
dGMP	deoxyguanosine 5'-monophosphate
DHAP	dihydroxyacetone phosphate
DIPEA	<i>N,N</i> -diisopropylethylamine
DMF	dimethylformamide
DMSO	dimethylsulfoxide
DMT	dimethoxytrityl
DN	double-negative
DNA	deoxyribonucleic acid
DOSY	diffusion ordered spectroscopy
DOTA	1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid
DPA	dipicolinic acid
DSA	dipolar shift anisotropy
DSF	differential scanning fluorimetry
DTPA	diethylenetriaminepentacetic acid
DTT	dithiothreitol
DTTA	diethylene-triamine-tetraacetate
DTTA-C3-yne	diethylene-triamine-tetraacetate propyl-1-yne
DTTA-C4-yne	diethylene-triamine-tetraacetate butyl-1-yne
ECAD12	first two domains of epithelial cadherin

EDTA	ethylenediaminetetraacetic acid
EHNA	erythro-9-(2-hydroxy-3-nonyl) adenine hydrochloride
EPR	electron paramagnetic resonance
ERp29	endoplasmic reticulum protein 29
ESI	electrospray ionization
EtOAc	ethyl acetate
EXSY	exchange spectroscopy
FACS	fluorescence-activated cell sorting
FBDD	fragment-based drug discovery
FDR	false discovery rate
FGE	formylglycine-generating enzyme
FGE-27	17-residue peptide substrate designed for the study of the FGE protein
FGEwt	wild type formylglycine-generating enzyme
FM-520	4-sulfamoyl- <i>N</i> -(4-(trifluoromethyl)benzyl) benzamide
fw _{hh}	full width at half height
G3P	glyceraldehyde 3-phosphate
GB1	B1 domain of protein G
GLO1	glyoxalase 1
GLOD4	glyoxalase domain containing protein 4
GM2, GM3	second and third discovered monosialic ganglioside
GO	gene-ontology
GO	glucose oxidase
GPCR	G protein-coupled receptor
GPS	global positioning system
Grb2 SH2	SH2 domain of growth factor receptor-bound protein 2
GSH	glutathione
GST	glutathione <i>S</i> -transferase
GXP	glutathione peroxidase
HATU	1-[bis(dimethylamino)methylene]-1 <i>H</i> -1,2,3-triazolo[4,5- <i>b</i>]pyridinium-3-oxidehexafluorophosphate
hCA II	human carbonic anhydrase II
HD	high definition
HeLa cell	immortal cell line derived from cervical cancer taken from Henriette Lacks
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
HetNOE	heteronuclear nuclear Overhauser effect
HiRes	high-resolution

HLA	human leukocyte antigen
HMQC	heteronuclear multiple-quantum correlation spectroscopy
HMBC	heteronuclear multiple-bond correlation spectroscopy
HOESY	heteronuclear Overhauser effect spectroscopy
HPLC	high-performance liquid chromatography
HPPK	ATP:2-amino-4-hydroxy-6-hydroxymethyl-7,8-dihydropteridine-6'-diphosphotransferase
HPRT1	hypoxanthine phosphoribosyltransferase 1
HR-220	Varian high resolution 220 MHz NMR spectrometer
HR-ESI-MS	high-resolution electrospray ionization mass spectrometry
HRMS	high-resolution mass spectrometry
Hsp90	heat shock protein 90
HSQC	heteronuclear single quantum correlation
HSQC IPAP	heteronuclear single quantum correlation in-phase/anti-phase
IBX	2-iodoxybenzoic acid
ICAM-1	intercellular adhesion molecule 1
IDA	iminodiacetic acid
IDP	intrinsically disordered protein
IFN- γ	Interferon- γ
IMPDH	inosine monophosphate dehydrogenases
INEPT	insensitive nuclei enhancement by polarization transfer
IPTG	isopropyl β -D-1-thiogalactopyranoside
IR	infrared
ITC	isothermal titration calorimetry
KEGG	Kyoto encyclopedia of genes and genomes
KO	knockout
LACC1	laccase domain containing 1
LacR	lac repressor
LBP	lanthanide binding peptide
LC	liquid chromatography
LCT	lanthanide chelating tag
LDAO	lauryldimethylamine oxide
LFA-3	lymphocyte function-associated antigen 3
LPMO	lytic polysaccharide monooxygenase
M1dG	3-((2 <i>R</i> ,4 <i>S</i> ,5 <i>R</i>)-4-hydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)pyrimido[1,2- <i>a</i>]purin-10(3 <i>H</i>)-one
M1G	pyrimido[1,2- <i>a</i>]purin-10(3 <i>H</i>)-one

M3Ade	8-(9 <i>H</i> -purin-6-yl)-2-oxa-8-azabicyclo[3.3.1]nona-3,6-diene-4,6-dicarbalddehyde
mAb	monoclonal antibody
MAIT cell	mucosal-associated invariant T-cell
MD	molecular dynamics
MDA	malondialdehyde
MFI	mean fluorescence intensity
MG	methylglyoxal
MMP-1	matrix metalloproteinase 1
MPA	mycophenolic acid
MR1	MHC-related molecule 1
MRI	magnetic resonance imaging
MR1T cell	MR1-restricted T-cell
MS	mass spectrometry
MSA	mercaptosuccinic acid
MSD	multiple sulfatase deficiency
MurD	UDP- <i>N</i> -acetylmuramoylalanine: <i>D</i> -glutamate ligase
NAC	<i>N</i> -acetylcysteine
NADPH	reduced form of nicotinamide adenine dinucleotide phosphate
NBS	<i>N</i> -bromosuccinimide
NMR	nuclear magnetic resonance
NMWL	nominal molecular weight limit
NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser effect spectroscopy
NTA	nitrilotriacetic acid
NUS	non-uniform sampling
OB-domain	DNA binding domain of single-stranded DNA binding protein
OPdA	<i>N</i> ⁶ -(3-Oxo-1-propenyl)-2'-deoxyadenosine
OPdC	<i>N</i> ⁴ -(3-Oxo-1-propenyl)-2'-deoxycytidine
Paz	domain of the proteins Piwi, Argonaut and Zwillie
PBMC	peripheral blood mononuclear cell
PCS	pseudocontact shift
PDB	protein data bank
PDE5A	cGMP-specific 3',5'-cyclic phosphodiesterase
PDZ domain	domain found in post-synaptic density protein, <i>Drosophila</i> disc large tumour suppressor and zonula occludens-1 protein
PDZ protein	family of proteins containing the PDZ domain

PHM	peptidylglycine- α -hydroxylating monooxygenase
pMMO	particulate methane monooxygenase
ppm	parts per million
PRE	paramagnetic relaxation enhancement
pSRII	phototaxis receptor sensory rhodopsin II
QCI	cryogenic quadruple resonance inverse NMR probe
RACS	residual anisotropic chemical shifts
RCM	ring-closing metathesis
RDC	residual dipolar coupling
RI	resolution of the identity
RMSD	root mean square deviation
RNA	ribonucleic acid
ROESY	rotating-frame Overhauser spectroscopy
ROS	reactive oxygen species
RSL-3	Ras-selective lethal small molecule 3
SAH domain	single α -helical domain
SAP	square antiprism
scFGE	FGE from <i>streptomyces coelicolor</i>
SD	standard deviation
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
sgRNA	single guide RNA
SH2 domain	Src homology 2 domain
SL	stem loop
SLC ^A	stem loop C ^A from the Moloney murine leukemia virus 5'-leader
SNAP-25	synaptosomal nerve-associated protein 25
SNARE	SNAP receptor
Src	proto-oncogene tyrosine-protein kinase (Src derived from sarcoma)
SrtA	sortase A
SSBP	single-stranded DNA binding protein
STRIDE	structural identification
Syt1	synaptotagmin 1
Subp2-R3H	R3H domain of Subp2
T3P	1-propanephosphonic acid anhydride
T4 Lys	T4 lysozyme
TAHA	triaminohexaacetate
TCEP	tris(2-carboxyethyl)phosphine
tcFGE	FGE from <i>thermomonospora curvata</i>

TCR	T-cell receptor
TFA	trifluoroacetic acid
THF	tetrahydrofuran
THP-1 cell	Tohoku hospital pediatrics-1 cell
TIL	tumor-infiltrating lymphocyte
TLC	thin-layer chromatography
TMSCI	trimethylsilyl chloride
TOCSY	total correlation spectroscopy
TOF	time-of-flight
TPI1	triosephosphate isomerase
Tris	2-amino-2-(hydroxymethyl)propane-1,3-diol
TROSY	transverse relaxation optimized spectroscopy
TSAP	twisted square antiprism
U1A	small nuclear ribonucleoprotein polypeptide A
UBA1	ubiquitin-like modifier activating enzyme 1
Ubi	ubiquitin
UPLC	ultra high performance liquid chromatography
UTR	unique tensor representation
UV	ultraviolet
VDAC	voltage-dependent anion channel
XAS	X-ray absorption spectroscopy
ZORA	zeroth order regular approximation

Table of Contents

1	Introduction	1
1.1	Nuclear magnetic resonance spectroscopy	1
1.2	Nuclear magnetic resonance spectroscopy of proteins.....	3
1.3	Nuclear magnetic resonance spectroscopy of ribonucleic acids	6
1.4	Paramagnetic nuclear magnetic resonance spectroscopy.....	8
1.5	Pseudocontact shift.....	10
1.6	Residual dipolar couplings.....	12
1.7	Paramagnetic relaxation enhancement	14
1.8	Lanthanide chelating tags.....	15
1.9	DOTA-based lanthanide chelating tags	17
1.10	Formylglycine-generating enzyme.....	20
1.11	Malondialdehyde Adducts of Nucleosides	22
2	Research goal	24
3	Part I	31
3.1	Conformationally locked lanthanide chelating tags for convenient pseudocontact shift protein nuclear magnetic resonance spectroscopy	31
3.2	A sterically overcrowded, isopropyl-substituted lanthanide chelating tag for protein pseudocontact shift NMR spectroscopy: Synthesis of its macrocyclic scaffold and benchmarking on ubiquitin S57C and hCA II S166C.....	61
3.3	P4T-DOTA - A lanthanide chelating tag combining a highly sterically overcrowded backbone with a reductively stable linker	115
3.4	A novel, rationally designed lanthanoid chelating tag delivers large paramagnetic structural restraints for biomolecular NMR	146
4	Part II	205
4.1	Design and application of lanthanide chelating tags for pseudocontact shift nuclear magnetic resonance spectroscopy on biomacromolecules.....	205
4.2	Application of paramagnetic lanthanoid chelating tags in NMR spectroscopy and their use for the localization of ligands within biomacromolecules.....	235
5	Part III	265

5.1	Induction of strong paramagnetic effects in RNA molecules by covalent lanthanide tagging for biomolecular NMR.....	265
6	Part IV	301
6.1	Structure of formylglycine-generating enzyme in complex with copper and a substrate reveals an acidic pocket for binding and activation of molecular oxygen	301
6.2	The molecular basis for multiple sulfatase deficiency caused by mutations in the outer calcium binding site of formylglycine-generating enzyme	331
7	Part V	362
7.1	T cell activation by nucleobase adduct-containing metabolites	362
8	Summary and Outlook.....	426

1 Introduction

1.1 Nuclear magnetic resonance spectroscopy

By today, NMR spectroscopy is one of the most important analytical techniques in chemistry and biology and is used in many different fields from organic synthesis to structural biology. Since practically all fields of chemistry deal with NMR-active nuclei as ^1H , ^{13}C , ^{15}N , ^{19}F , ^{31}P and potentially also ^{11}B , the development of a large variety of NMR methodologies allows to characterize the molecular architecture and purity of chemical compounds. In chemistry, NMR is nowadays mainly used as a tool for structure elucidation and confirmation. However, NMR is not only able to be used for purely structural studies, also thermodynamic and kinetic barriers, flexible parts of proteins or the binding of ligands can be identified and studied.¹⁻⁷ The different homo- and heteronuclear NMR experiments including pulse sequences for the measurement of through-space and through-bond correlations between nuclei and techniques suitable for complex structural studies, were and are crucial to the rational design of drugs and understanding of structure and function of biomacromolecules, e.g. proteins, nucleic acids and carbohydrates. The significant technical developments of NMR spectroscopy over the last decades, including the superconducting, helium cooled magnets (Figure 1) as well as highly sensitive, cryogenic probe heads with a very low background noise enabled the nowadays world-wide and indispensable application of NMR experiments.

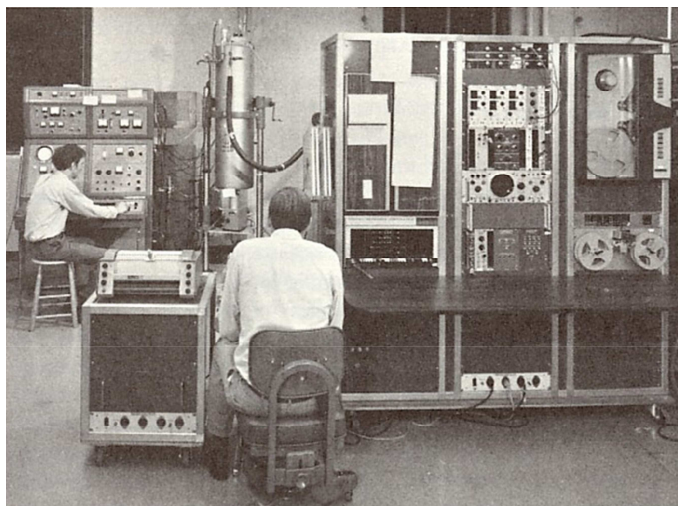
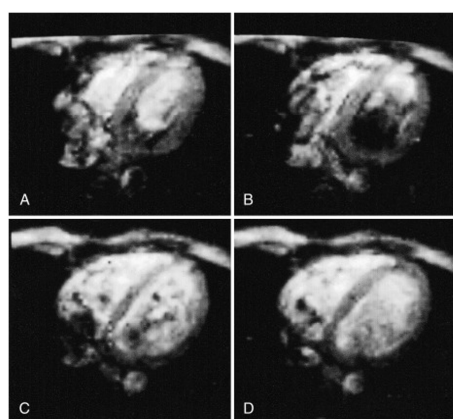


Figure 1: A Varian HR-220, one of the first NMR spectrometers with a superconducting coil in a dewar of liquid helium placed in a liquid nitrogen reservoir.⁸

The journey of the development of NMR spectroscopy started in 1938, when Isidor Rabi first described the phenomenon of NMR measured in molecular beams by giving experimental proof that atoms in an atom beam through a constant magnetic field can be brought into resonance with an external magnetic field.⁹⁻¹¹ In 1944, Isidor Rabi was then awarded the Nobel

prize in physics for his discovery.¹² Felix Bloch and Edward Purcell further investigated the newly developed technique by observing the energy transfer from the external magnetic field to the spins of the nuclei and the alternating current induced by the precessing dipoles in a conducting coil. The discoveries of Felix Bloch and Edward Purcell allowed to expand the NMR technique for the measurement of liquids and solids and the researchers shared the Nobel prize in physics in 1952.¹³⁻¹⁵ Thorough studies of the Larmor precession and the induced current by the dipoles, free induction decay and longitudinal and transverse relaxation were conducted by Erwin Hahn.¹⁶⁻²⁰

Based on the developments for NMR, Peter Mansfield and Paul Lauterbur exploited the signals of the spins of the nuclei to develop a 3-dimensional method called magnetic resonance imaging (MRI).²¹⁻²² By the application of strong gradients onto a homogeneous static magnetic field, the resonance frequency of the nuclei within the magnetic field depends on the position of these nuclei. In combination with the relaxation time, high resolution images for medical diagnoses can be obtained and the MRI technique is today indispensable in hospitals and medical centers all around the world (Figure 2).



*Figure 2: MRI imaging of the human heart with a spatial resolution better than 2 mm.^{21, 23} Reprinted with permission from P. Mansfield, *Angew. Chem. Int. Ed.*, 2004, 43, 5456-5464.²¹ Original figure published in: M. J. Stehling, A. M. Howseman, R. J. Ordidge, B. Chapman, R. Turner, R. Coxon, P. Glover, P. Mansfield, R. E. Coupland, *Radiology*, 1989, 170, 257-263. Copyright 1989 Radiological Society of North America.²³*

By introduction and combination of a varying delay and Fourier-transformation in two dimensions in so-called two-dimensional NMR experiments, two nuclei can be correlated and thus more detailed information about the structure of a studied molecule can be obtained. Through the developments and the practical implementation of 2D-NMR experiments by Richard Ernst, Ray Freeman and other contributors in the 1970's and 1980's, experiments as e.g. COSY, NOESY, TOCSY, DOSY, HMQC, HMBC and composite decoupling pulses became available and are today the foundation for the structure elucidation of small and

medium sized molecules.²⁴⁻³² Based on these two-dimensional homo- and heteronuclear NMR experiments, through-space and through-bond correlations within the same and between different sorts of NMR-active nuclei can be established and analyzed. In order to access proteins by NMR, further developments considering sensitivity and number of the involved nuclei (multidimensional NMR experiments) were implemented.

1.2 Nuclear magnetic resonance spectroscopy of proteins

In research fields as drug design, structural biology and biochemistry, the investigation of the solution structure of biomacromolecules, e.g. proteins or nucleic acids, as well as their functions, dynamics and complexes, is crucial for the understanding of the fundamental processes of life.³³⁻³⁴ X-ray crystallography is able to give insights into the static picture of a protein in crystalline form,³⁵ however, due to the high interest into the solution structure and dynamics of proteins, new measurement techniques had to be developed. Solution NMR stands out as a unique tool in this context, due to its ability to provide solution structures in atomic details, as well as the opportunity to adjust the experimental conditions in order to mimic different physiological environments.³⁴ Besides being able to adjust important parameters as pH, salt concentration, concentration of the biomacromolecule, additives as reducing agents and ligands, also *in cell* NMR measurement of proteins was successfully demonstrated.³⁶⁻³⁹ The obtained structural information can be compared to the crystal structure gained by X-ray crystallography in order to identify dynamic or disordered regions in proteins. The measurement of protein ligand binding by NMR offers the possibility to conveniently exchange the ligand for subsequent experiments as well as the recording of competition experiments.

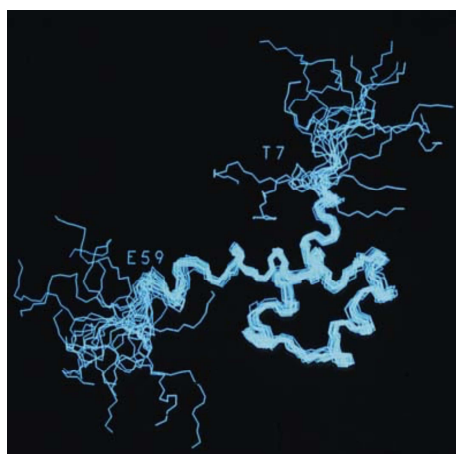


Figure 3: NMR structure of the Antennapedia homeodomain displaying for the polypeptide segment a highly defined structure in the middle part and disordered ends of the chains.³⁴

In order to obtain structural information from protein NMR, multi-dimensional NMR experiments were developed in order to establish the through-bond correlations between neighboring nuclei. Subsequently, NOESY experiments and computational tools were combined for structure calculations. Many of these needed experiments and techniques were developed and studied by Kurt Wüthrich, a Swiss protein NMR spectroscopist that was awarded the Nobel prize in 2002 for his development of NMR methods to determine the 3-dimensional structure of biological macromolecules in solution (Figure 3).³⁴ Wüthrich and coworkers reported in 1984 the first structure of a globular protein in solution state solely solved by NMR spectroscopy.⁴⁰⁻
⁴¹ Crucial features of protein NMR experiments are the various techniques for water signal suppression⁴²⁻⁴³ and an initial INEPT step in the experiments in order to enhance the sensitivity of the experiment by starting on the N-H proton.⁴⁴ Subsequently, different magnetization pathways can be followed and detected in order to establish the correlations between neighboring nuclei. The development of a large variety of two-dimensional and triple-resonance NMR experiments for protein NMR led to the astonishing toolbox available today for both the assignment of protein resonances and the following structure determination. Based on a backbone assignment of a protein by 3-dimensional experiments, developed by Lewis E. Kay, Ad Bax and co-workers, as e.g. HNC0, HNCA, HNCACO and HNCOCA, information about amino acid sequence and dihedral angles can be obtained.⁴⁵⁻⁴⁷ Upon a full assignment of the protein residues including the sidechains, structural restraints from the obtained NOESY correlations can then be used to generate a 3-dimensional model of the structure of the protein. However, due to the limitations of NOESY correlations, as e.g. cross-relaxation effects with nearby nuclei and the distance restriction to around 5-8 Å,³⁴ alternative methods for the measurement of long-range structural restraints by paramagnetic NMR become highly valuable tools. For larger proteins, perdeuteration of the protein can be used in order to decrease the cross-relaxation of the measured amide protons with sidechain protons.⁴⁸ The absence of the majority of proton nuclei within the protein scaffold leads to a slower relaxation of the measured amide protons. In combination with the TROSY effect that was first described by Pervushin et al.,⁴⁹ recording of multidimensional NMR experiments of large proteins and their complexes became feasible (Figure 4).⁴⁹⁻⁵²

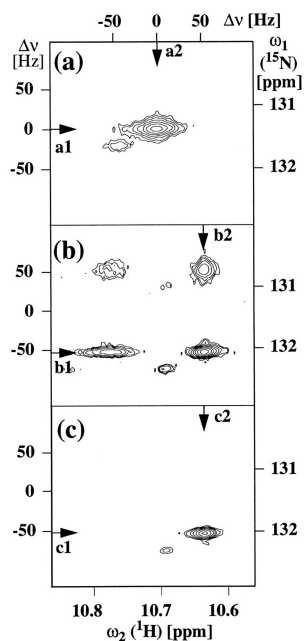


Figure 4: a) Broad-band decoupled ^1H - ^{15}N HSQC. b) In both dimensions non-decoupled ^1H - ^{15}N HSQC. c) TROSY-type ^1H - ^{15}N HSQC selecting the sharp component of the signal (Copyright 1997 National Academy of Sciences).⁴⁹

Modern methods in protein NMR spectroscopy use relaxation dispersion for studying invisible, low-populated states or folding intermediates.⁵³ Furthermore, the development of the SOFAST experiment⁵⁴ with selective excitation bandwidths enables fast repetition rates due to the increased T_1 relaxation. The SOFAST methodology and non-uniform sampling (NUS)⁵⁵, which reduces the required increments in the indirect dimension strongly, reduced the needed measurement time for acquisition of multidimensional NMR spectra. For study of very large molecular entities, as e.g. G protein-coupled receptors (GPCRs) or antibodies in solution, novel methods such as the XL-ALSOFAST experiment⁵⁶ with delayed decoupling may be applied to boost the sensitivity of the experiment. Apart from the expansion of the NMR toolbox to proteins with a molecular weight in the mega Dalton range, the study of the function of chaperones⁵⁷ and liquid-liquid phase separation⁵⁸ constitute important current topics in the field of biomolecular NMR.

Protein NMR spectroscopy is not only suited to obtain information about structure and dynamics of a single protein, but can also be extended to protein-protein,⁵⁹ protein-ligand⁶⁰⁻⁶¹ and protein-nucleic acid complexes.⁶² Methods as CSP and NOESY allow the identification of docking sites and measurement of distances between two molecular entities. However, potentially ambiguous chemical shift changes in CSP experiments and the distance limitation of NOESY restraints to about 5-8 Å render paramagnetic NMR spectroscopy as a highly valuable tool to localize and study molecular structures over increased distances of up to 60 Å.⁶³

1.3 Nuclear magnetic resonance spectroscopy of ribonucleic acids

While the NMR methodologies for the study of proteins and their complexes are nowadays well established, the structure elucidation of small to medium RNA strands as well as the characterization of their dynamics and complexes becomes increasingly important.⁶⁴⁻⁶⁸ In contrast to DNA, which is predominantly present in helical structures, it became evident over the last decades that RNA molecules exhibit a rich diversity of solution state structures. Furthermore, RNA molecules take part in the regulation machinery of important processes in a cell, e.g. the regulation of transcription by small RNA molecules⁶⁹ or the regulation of translation by RNA secondary structure elements.⁷⁰ Most interestingly, RNA molecules are at the starting point to become the next generation of drug targets and their structural diversity offers enormous potential for future targeting with drug molecules.⁷¹ Since NMR spectroscopy is able to reveal the solution characteristics of RNA, RNA-RNA and RNA-ligand complexes at atomic resolution, NMR spectroscopy of RNA molecules is an increasingly important subject. Upon completion of a resonance assignment of a targeted RNA strand, NMR spectroscopy can be used to yield information about:^{66, 68}

- The base-pairing pattern by chemical shifts of imino protons, HNN-COSY experiments and evaluation of NOEs
- The formation and equilibria of structure elements as e.g. hairpins, bulges and single-strand/duplex structures
- The interaction of RNA strands with other RNA molecules, proteins, ligands or metal ions.

The production of an isotope-labeled RNA strand is either conducted by enzymatic methods or solid-phase synthesis, whereby the enzymatic method circumvents the need for expensive precursors as in oligonucleotide synthesis. However, the enzymatic production of RNA strands can lead to inhomogeneous products by adding one or two more nucleobases at the end of the sequence that are not encoded in the plasmid DNA.⁶⁸ The solid-phase synthesis omits this issue and offers the possibility to include varying artificial nucleotides and site-specific labeling of nucleotides.⁷²

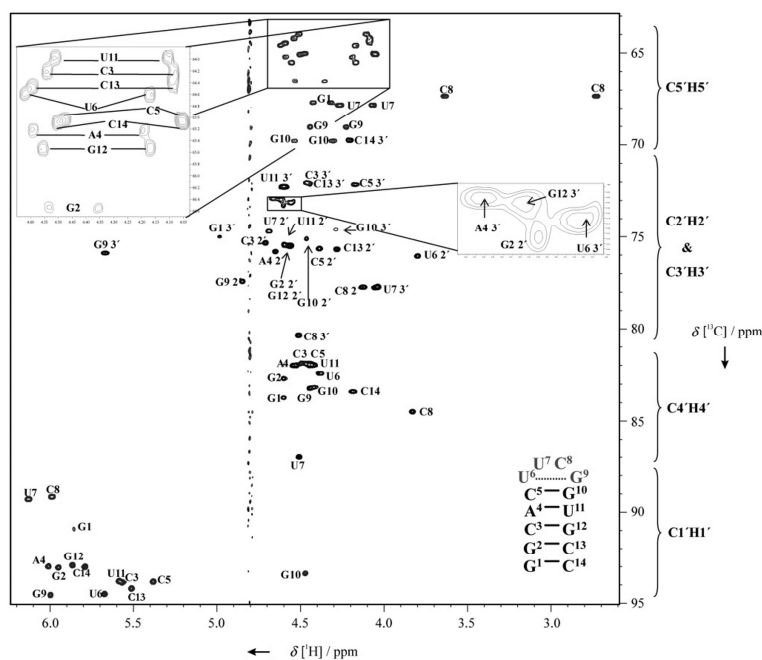


Figure 5: ^1H - ^{13}C -HSQC spectrum of the UUCG tetraloop recorded at 600 MHz proton frequency. The structure is indicated in the lower right corner of the spectrum.⁶⁸

The sequential assignment of the imino protons and specific protons of the deoxyribose moieties within an RNA molecule can be obtained via a NOESY walk. Subsequently, homonuclear, through-bond NMR experiments, e.g. TOCSY, allow the assignment of yet unassigned protons within individual spin systems. Heteronuclear correlation experiments complement the assignment for ^{13}C , ^{15}N and ^{31}P nuclei (Figure 5) and a thorough analysis of coupling constants and chemical shifts provides insights into the base-pairing as well as adopted secondary structural elements and therefore into the local structure of the investigated RNA molecule. Furthermore, McCallum et al. showed that it is possible to gain insights into the global structure of RNA molecules by use of RDCs induced by filamentous phages.⁷³ The presence of filamentous phages within the sample leads to an anisotropic tumbling of the RNA molecules and enables thereby the measurement of RDCs.⁷³ While CSPs and NOEs offer the possibility to investigate binding of ligands, metal ions that are invisible in NMR experiments as e.g. Mg^{2+} can be substituted by paramagnetic Mn^{2+} . The substitution of diamagnetic Mg^{2+} with paramagnetic Mn^{2+} leads to broadening of signals due to paramagnetic relaxation effects (PREs) induced by the unpaired electrons of the paramagnetic substitute. The increase of relaxation rates in the paramagnetic species when compared to the diamagnetic reference can then be used to identify binding sites for ions that might be crucial for stabilization of secondary structure elements within the RNA strand.⁶⁶ For the study of interactions of RNA strands with proteins, HSQC experiments can be acquired for both the protein and the RNA molecule in the complex to obtain complementary information about the binding interface.⁶⁸

While it is straightforward nowadays to investigate the local conformations of an RNA strand, the lack of precise experimental long-range restraints severely impedes the characterization of the arrangement of individual secondary structure elements and domains within the RNA molecule. Paramagnetic NMR spectroscopy and in particular measurement of PCSs and RDCs offer the possibility to acquire such urgently needed long-range restraints^{59, 74-75} and thus constitute a highly auspicious methodology for the characterization of the global structure of RNA strands in solution. The use of PCS restraints for the elucidation of the binding interface of RNA complexes with ligands and proteins can complement CSPs, which potentially could lead to ambiguous results, as well as NOESY restraints, which are only detected over a short distance range of 5-8 Å. Today, the toolbox for paramagnetic NMR on RNA molecules is severely underdeveloped and advances in the design of LCTs and suitable linkers are strongly encouraged. While the use of LCTs in protein NMR is well established but still far away from being perfect, the characterization of RNA solution structures by the application of paramagnetic lanthanide complexes remains an entirely open field for future developments.

1.4 Paramagnetic nuclear magnetic resonance spectroscopy

The distance limitation of the NOESY effect to 5-8 Å and potentially ambiguous CSP experiments render paramagnetic effects as a highly valuable tool to identify and measure distance restraints for the study of biomacromolecules.^{59, 74-96} Paramagnetic effects observed in NMR include paramagnetic shifts (contact and pseudocontact shifts),^{74, 76-77, 79-80, 82-83, 88, 97-100} paramagnetic relaxation enhancement^{90-92, 101-106} and residual dipolar couplings.^{59, 77, 88, 107-110} The different effects associated with the introduction of paramagnetic centers yield complementary restraints and, therefore, their combination represents a powerful tool for the structure determination of biomacromolecules (Figure 6). Whereas contact shifts, that are through-bond effects, can be neglected for studies of biomacromolecules, PCSs yield structural restraints up to distance of 60 Å due to their favorable distance dependence of r^{-3} and the 3D information that is contained in the induced shift of the signals.^{63, 74} PREs offer the opportunity to obtain information about the distance of a nucleus from the paramagnetic center by means of line broadening and the associated decrease in signal intensity. The distance dependence of PRE effects (r^{-6}) is less favorable than the distance dependence of PCSs (r^{-3}) and no angular information is obtained, however, the magnitude of PRE effects suffers less from motional averaging. RDCs offer information about the orientation of the bond vector of 2 coupled nuclei in space and can for example be combined with a Rosetta calculation to derive a *de novo* structure of a protein in solution and inside living cells. In order to observe RDCs, partial alignment of the protein is needed. This can be obtained by either the introduction of a

metal center with an anisotropic electron distribution or by use of an alignment medium as e.g. an acrylamide gel.¹¹¹⁻¹¹²

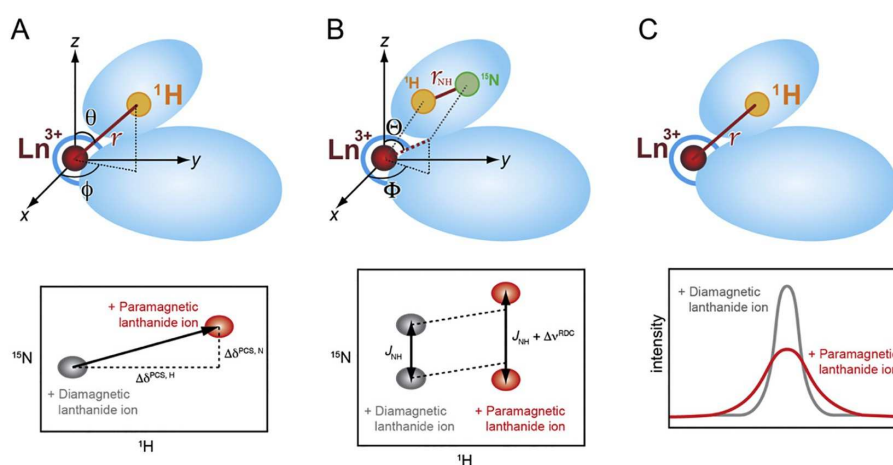


Figure 6: The introduction of a paramagnetic lanthanide ion leads to significant effects in NMR spectra: A) Localization of measured spins using PCSs and PCSs observed in a ^1H - ^{15}N HSQC experiment. B) Orientation of the amide bond vector revealed by RDCs and RDCs observed in a ^1H - ^{15}N HSQC IPAP experiment. C) Determination of the distance of a measured spin from the paramagnetic center by PRE and signal decrease/line broadening caused by PRE.¹¹³

The theoretical background of the NMR shifts induced by paramagnetic lanthanides was established in the 1960s by Bleaney.¹¹⁴⁻¹¹⁵ However, recent critics and findings deliver a new perspective on the validity of Bleaney's theory concerning the magnetic properties of lanthanide complexes used in paramagnetic NMR spectroscopy.¹¹⁶⁻¹¹⁷ The very first applications of paramagnetic NMR include the europium shift reagents that increase the chemical shift range in order to resolve signal overlap and to allow the determination of enantiomeric ratios by formation of diastereomeric complexes with chiral molecules *in situ*.¹¹⁸⁻¹²¹ After the successful exploitation of paramagnetic effects of metal ions contained in metalloproteins by Bertini and coworkers,^{90, 95} and determination of the associated paramagnetic susceptibility tensors, the next step in the development of paramagnetic NMR constituted lanthanide binding peptides (LBPs) mimicking the zinc finger motif and first steps towards single- and double-armed lanthanide binding tags that can be attached to a protein.¹²² Nowadays, paramagnetic NMR has found wide-spread applications in structural biology, such as the refinement of X-ray structures,¹²³ investigation of the solution dynamics of proteins,¹²⁴ localization of ligands within proteins (Figure 7),^{94, 125-129} study of protein-protein complexes,¹³⁰ enhancement of signal dispersion in crowded HSQC spectra¹³¹ and analysis of large protein complexes.⁸⁵

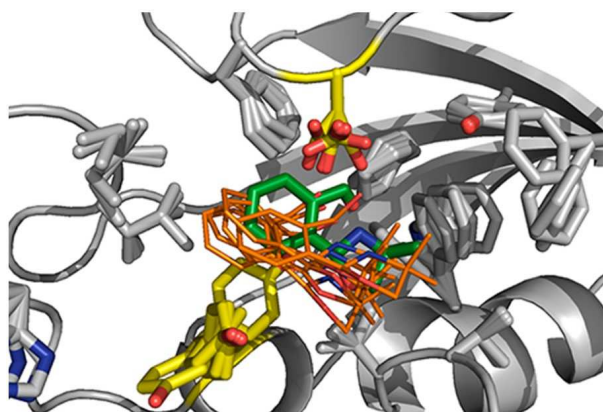


Figure 7: GPS-approach for the localization of a ligand within a protein by use of PCS restraints (green: localization of the ligand by NOE restraints, orange: five obtained structures with the least violation of experimental PCS restraints). Reprinted with permission from J.-Y. Guan, P. H. J. Keizers, W.-M. Liu, F. Löhr, S. P. Skinner, E. A. Heeneman, H. Schwalbe, M. Ubbink, G. Siegal, *J. Am. Chem. Soc.*, 2013, 135, 5859-5868. Copyright 2013 American Chemical Society.¹²⁷

Besides the application of paramagnetic effects in protein NMR spectroscopy, also basic research studies on solvent dependence as well as temperature and pH effects on the magnetic anisotropy induced by paramagnetic molecules alone and when attached to proteins were performed.^{99, 132}

Furthermore, computational approaches as e.g. DFT calculations allow to predict the structure and the equilibrium of conformational species of lanthanide complexes *in vacuo* and in solution as well as the anisotropy tensor for EPR studies.¹³³⁻¹³⁹ The computational approach allows by more elaborate predictions in future to decrease the extensive work at the bench associated with the development of new LCTs.

1.5 Pseudocontact shift

The through-space effect on the shift of signals in NMR associated with the introduction of a paramagnetic center is termed pseudocontact shift. The PCS arises from a contribution of the unpaired electrons of the paramagnetic center to the effective magnetic field that a specific nucleus experiences. This shift contribution is dependent on the distance and angular positioning of the nucleus with respect to the paramagnetic center. Considering this 3D spatial information and the distance dependence of the effect of r^{-3} , highly valuable structural restraints of the investigated biomacromolecule in vicinity of the paramagnetic center can be obtained. The PCS that a nucleus displays can be calculated from the following equation that can be expressed either in polar (Equation I) or cartesian coordinates (Equation II):¹⁴⁰

- (I) Pseudocontact shift (polar coordinates) $\delta^{PCS} = \frac{1}{12\pi r^3} \left[\Delta\chi_{ax} (3\cos^2\theta - 1) + \frac{3}{2} \Delta\chi_{rh} \sin^2\theta \cos 2\phi \right]$
- (II) Pseudocontact shift (cartesian coordinates) $\delta_i^{PCS}(x, y, z) = \frac{1}{12\pi r^3} \left[\Delta\chi_{ax} \frac{2z^2 - x^2 - y^2}{r^2} + \frac{3}{2} \Delta\chi_{rh} \frac{x^2 - y^2}{r^2} \right]$

In order to describe the anisotropy tensor associated with a paramagnetic center, e.g. of an LCT when attached to a biomacromolecule, 8 parameters are needed:

- The x, y and z position of the metal center
- The three Euler angles α , β and γ that describe the orientation of the anisotropy tensor with regard to the biomacromolecule
- The axial anisotropy parameter ($\Delta\chi_{ax}$)
- The rhombic anisotropy parameter ($\Delta\chi_{rh}$)

As depicted in Figure 8, the anisotropy of the χ -tensor is usually displayed as isosurfaces with a given PCS value. For a pronounced high axial/rhombic ratio, a shape resembling a d_{z^2} orbital is obtained, whereas for a small axial/rhombic ratio a shape resembling to a 45° rotated d_{xy} orbital is obtained.

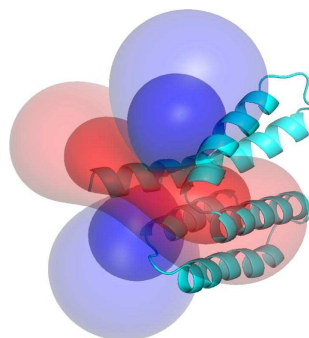


Figure 8: PCS isosurfaces of a Tb^{3+} complex attached to the ERp29 C-terminal domain (inner layer: 0.7 ppm, outer layer: 0.2 ppm).⁷⁴

In order to study proteins by PCS NMR, usually HSQC or HSQC-TROSY experiments with a diamagnetic and a paramagnetic sample are recorded. Based on the PCSs observed for the paramagnetic sample when compared to the diamagnetic sample, the anisotropy parameters can be fitted and an initial structure of a biomacromolecule can be refined.¹⁴¹ Due to the close positioning of correlating nuclei in through-bond, two-dimensional NMR experiments, the two coupling nuclei most often experience the same PCS contribution (in ppm). An exception from this result is observed, when the coupling nuclei are located in very close proximity to the anisotropic, paramagnetic center. Due to the equal PCS contribution in both dimensions, an overlay of the diamagnetic and paramagnetic spectrum leads usually to a very specific shift pattern in 2D spectra. The recorded spectra are then iteratively assigned in combination with

back-calculating the theoretical PCSs from the initial structure. PCSs enable thereby a large variety of intriguing applications from structure refinement to localization of ligands.^{77, 94-95, 125-129} Deviations from the back-calculated shift using the solid-state structure can be used to determine flexible regions in the protein. Furthermore, the combination of PCS NMR spectroscopy in combination with Rosetta allows to obtain an accurate representation of the protein's solution structure.¹⁴² The required input for the structure determination of a protein is then reduced to PCS data for all amide protons and the sequence. The combination of the Rosetta database that contains structural data for short peptide segments and the experimental restraints given by the PCS dataset enables a *de novo* structure determination of a protein.⁷⁷ Besides the use of PCSs as structural restraints for the analysis of protein structures, they can also be exploited for the localization of ligand binding pockets.^{94, 125-129, 143} By taking into account the PCS data from at least three different tagging sites, a ligand bound to a protein can be localized based solely on its PCS. This localization method was already demonstrated for ¹H PCSs for weakly bound ligands and was further developed in our group by K. Zimmermann et al. for ¹⁹F PCS NMR spectroscopy with high affinity ligands.^{127, 143}

1.6 Residual dipolar couplings

RDCs can be observed due to partial alignment of a molecule that leads to an incomplete averaging of dipolar couplings.¹¹¹ Under standard conditions, the protein tumbles in solution isotropically resulting in the averaging of dipolar interactions. If dipolar couplings are required to be observed, partial alignment is therefore required. A partial alignment of a biomacromolecule in solution can be achieved by either introducing an anisotropic, paramagnetic center attached to the protein, bicelles and filamentous phages leading to a field-dependent alignment,^{59, 77, 88, 107-111, 144-145} or using an alignment medium such as polyacrylamide leading to a field-independent alignment.^{111, 145} RDCs are a valuable source of long-range structural restraints and can be used in order to give insights into specific biological problems, as e.g. the structural properties of a long helix in a protein, the determination of the relative orientation or exact angle between two helices in proteins and the orientation of two subunits of a protein or protein-protein complex.^{59, 108, 146} Based on RDC and PCS data, the anisotropy parameters can be fitted and, subsequently, a refined structure can be obtained.¹¹² In principle, the orientation of every N-H bond vector can be obtained by RDCs, complementary to PCS experiments where the 3D positioning of individual nuclei is determined. Therefore, RDC data in combination with PCS restraints leads to a realistic picture of the solution structure of the investigated biomacromolecule in solution.⁷⁷ In contrast to PCSs (r^3) or PREs (r^6), RDCs are in a first approximation distance independent (Equation III).

(III) Dipolar coupling for two spins i and j $D_{ij} = k_{ij} \left[A_a (3 \cos^2 \theta - 1) + \frac{3}{2} A_r \sin^2 \theta \cos(2\phi) \right]$

with $k_{ij} = -\frac{\gamma_i \gamma_j \mu_0 h}{16\pi^3 r_{ij}^3}$

DOTA-based LCTs as e.g. DOTA-M8-(8S)-SSPy, an LCT developed by D. Häussinger, can yield such valuable structural restraints.⁸¹ Furthermore, as recently demonstrated by Müntener et al., using a DOTA-based tag bearing a reduction stable pyridine linker, RDCs of up to 20 Hz at 293 K were observed using *in vitro* and *in cell* NMR experiments (Figure 9).⁷⁷ These orientational restraints were then used to perform a *de novo* structure determination for GB1 in combination with the Rosetta database. This is particularly interesting for proteins without the possibility to record an X-ray structure or when deviations for the structure in solution from the crystal structure are to be expected.

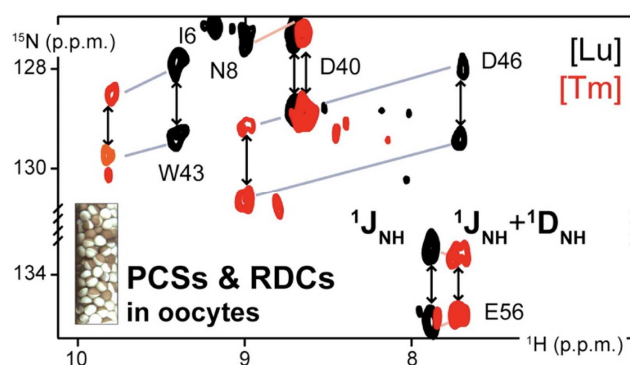


Figure 9: Measurement of PCSs and RDCs with ^1H - ^{15}N HSQC IPAP experiments on GB1 with an attached Ln-M7Py LCT in *Xenopus laevis* oocytes. Adapted with permission from T. Müntener, D. Häussinger, P. Selenko, F.-X. Theillet, *J. Phys. Chem. Lett.*, 2016, 7, 2821-2825. Copyright 2016 American Chemical Society.⁷⁷

The combination of RDCs with further diamagnetic and paramagnetic NMR restraints and other methods as e.g. small-angle X-ray scattering can lead to highly refined and realistic solution structures of large assemblies of biomacromolecules.¹⁴⁷ Interestingly, the inherent flexibility of a biomacromolecule can lead to an averaging of RDCs that are induced by an LCT, which is ligated to a specific position in the biomacromolecule. This property of RDCs that are induced by an LCT, can be thoroughly exploited for the characterization of the stiffness of long helices or the relative movement of two domains in a biomacromolecule.^{107, 148} Furthermore, RDCs were used to investigate the helix curvature, stoichiometry and domain orientation of homomultimeric nucleic acid complexes¹⁴⁹ as well as the determination of the stereochemistry of natural products.¹⁵⁰

1.7 Paramagnetic relaxation enhancement

PRE leads to line broadening and hence to a decrease in signal intensity of spins close to a paramagnetic center due to an enhancement of relaxation rates.^{91, 140} The distance dependence of the effect scales with r^6 and with strongly paramagnetic LCTs, the effect can be observed over a distance of up to 35 Å.⁹¹ Since PREs are convenient to analyze and less prone to be affected by motional averaging when compared to PCSs and RDCs, they are a broadly applied source of structural restraints in structural biology despite the fact that PCSs and RDCs have a more favorable distance dependency.⁷⁴ PREs are commonly induced by application of spin labels containing unpaired electrons, e.g. nitroxide tags or gadolinium(III) incorporated in suitable chelators.⁹¹ Whereas for nitroxide and gadolinium(III) spin labels the Curie-spin relaxation is negligible and the Solomon-Bloembergen type relaxation dominates (Equation IV), the Curie-spin relaxation (Equation V) constitutes the major component for lanthanide ions with an anisotropic g-tensor and a short electron relaxation time (all except lanthanum(III), gadolinium(III) and lutetium(III)).⁹¹

$$(IV) \quad \text{Solomon-Bloembergen} \quad \Gamma_2 = \frac{1}{15} \left(\frac{\mu_0}{4\pi} \right)^2 \gamma_I^2 g^2 \mu_B^2 S(S+1) \{4J_{SB}(0) + 3J_{SB}(\omega_I)\}$$

$$\text{with} \quad J_{SB}(\omega) = r^{-6} \frac{\tau_c}{1+(\omega\tau_c)^2}$$

$$(V) \quad \text{Curie-Spin} \quad \Gamma_2 = \frac{1}{5} \left(\frac{\mu_0}{4\pi} \right)^2 \frac{\omega_I^2 g^4 \mu_B^4 S^2 (S+1)^2}{(3k_B T)^2 r^6} \left(4\tau_r + \frac{3\tau_r}{1+(\omega_I\tau_r)^2} - 4\tau_c - \frac{3\tau_c}{1+(\omega_I\tau_c)^2} \right)$$

Out of the series of the lanthanide metals, gadolinium(III) with its isotropic electron distribution shows the strongest PRE effect, whereas no PCSs can be observed.

Besides the attachment of gadolinium(III) tags to biomacromolecules, the effect of the paramagnetic metal ion on the relaxation rates in its surroundings can be thoroughly exploited for biomedical purposes. Gadolinium-DOTA complexes are widely used in MRI as well as in radiopharmacy.¹⁵¹⁻¹⁵² Basically, two different mechanisms of relaxation are observed for the interaction of nuclear spins with the unpaired electrons of the gadolinium complexes. Whereas for water molecules that are directly coordinated to gadolinium(III)-based MRI contrast agents an inner-sphere mechanism governs the relaxation, the outer-sphere relaxation mechanism affects all nearby water molecules around the gadolinium complex (Figure 10).¹⁵³ In MRI contrast agents, both mechanisms contribute substantially to the relaxation enhancement of nearby protons.¹⁵³ When inducing PREs with paramagnetic tags, the outer-shell mechanism affects the protons of the protein or nucleic acid under investigation.

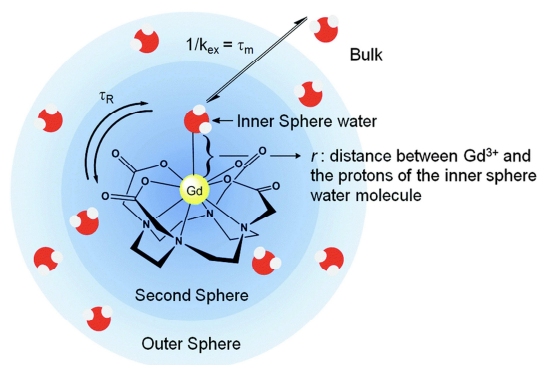


Figure 10: Inner- and outer-sphere relaxation on Gd-DOTA complexes. Reproduced from Ref. ¹⁵³ with permission from The Royal Society of Chemistry.¹⁵³

Spin labels, e.g. nitroxides or Mn^{2+}/Gd^{3+} metal chelates, were extensively used and successfully introduced into proteins, nucleic acids and their complexes in order to get structural restraints by PRE.^{91, 102, 154-156} Rigidified DOTA-based LCTs are more challenging to obtain. However, stronger PREs can potentially be obtained by usage of these LCTs for PRE studies of biomacromolecules and therefore the further development of the tags is highly desirable.

In contrast to PCS restraints, no angular information can be obtained with PRE. However, for example for the localization of ligands in metal binding proteins, combining PREs observed on the ligand molecules together with a known structure displays still a significant advantage over NOESY restraints, since in order to obtain useful NOESY restraints the whole protein sidechains have to be assigned. Both PREs and PCSs can be combined to localize ligands on a proteins surface in fragment based drug discovery.¹²⁸

In NMR studies of the applied LCTs themselves, the extremely fast T_2 relaxation provides challenges, since e.g. for gadolinium-, thulium- and dysprosium complexes, no two-dimensional NMR experiments can be successfully implemented.¹⁵⁷ For studies of these strongly paramagnetic LCTs, e.g. DFT calculations can fill the gap and provide insights into the structure and conformation.^{133, 135, 158}

1.8 Lanthanide chelating tags

PCS NMR spectroscopy using lanthanides started with the determination of nucleotide conformations in solution investigated with lanthanide induced shifts, that were observed in NMR experiments.¹⁵⁹ Later, Bertini and co-workers showed that intrinsic metal binding sites in metalloproteins can be exploited for the acquisition of structural restraints by use of anisotropic, paramagnetic metal ions.¹⁶⁰ Upon successful substitution of the calcium ion, paramagnetic effects, e.g. PCS and RDC, can be measured directly, as shown by Bertini et al. for Calbindin D_{9k}.¹⁶⁰ However, this approach is only applicable to proteins that contain a metal binding site.

Therefore, approaches that introduce artificial metal binding sites or chelators that are attached to the biomacromolecule of interest were urgently needed.

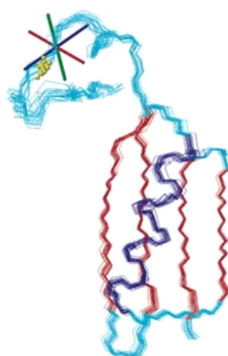


Figure 11: Solution state structure of GB1 refined with PCS data recorded upon the attachment of a two-point anchored LBP (20 lowest-energy structures, yellow dots: metal position, cross: tensor orientation).¹⁶¹

An important milestone for the introduction of paramagnetic metal ions into non-metalloproteins was the development of LBPs (Figure 11). The peptides mimic EF hand and zinc finger motifs in order to introduce lanthanide ions into proteins with no natural metal binding site.¹⁶²⁻¹⁶⁵ LBPs are advantageous in regard to tagging efficiency since coexpression of the LBP always leads to complete “ligation yields”. However, a severe drawback is that the incorporation of the LBP via coexpression is only possible on the C- or N-terminus of the investigated protein and furthermore, labeling of the LBP leads to increased complexity of protein NMR spectra. In order to gain more flexibility regarding the choice of the attachment site, LBPs that are ligated to cysteine residues were developed.¹⁶⁶⁻¹⁶⁷ After the tagging reaction, the protein can be purified by ion exchange chromatography. However, LBPs do not provide optimal results during the application process due to the restriction to the C- and N-terminus as attachment site and the remaining flexibility of cysteine-bound LBPs.

In order to circumvent the drawbacks associated with LBPs, synthetic LCTs were developed. Based on chelators as e.g. DTPA¹⁶⁸ and EDTA,^{87, 169} synthetically produced metal binding sites can be provided. The tags have to be attached to a specific ligation site on the protein, whereas the thiol group of cysteine stands out as an ideal target due to the low abundance in protein sequences and the opportunity for a chemoselective modification by LCTs. For the establishment of a rigid tagging site in a well-folded region of the protein, typically cysteine residues in alpha-helices or beta-sheets are chosen as attachment point. In order to ligate LCTs to cysteine residues in proteins, suitable activator moieties had to be developed. Very practical systems constitute the pyridinethione as well as methyl- and phenylsulfonyl activated linkers.⁷⁴ Recently, also artificial amino acids were introduced by Otting et al. that can be tagged in a bioorthogonal fashion.⁷⁹

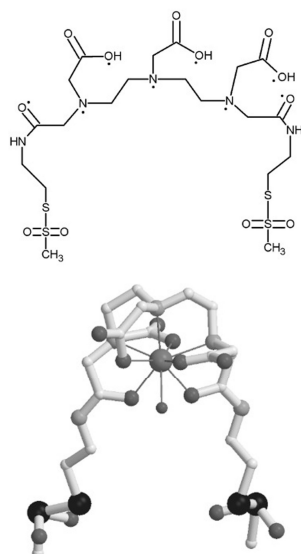


Figure 12: Structure of the DTPA-based two-point anchored LCT CLaNP (dots: coordinating atoms).¹⁶⁸

When the linker is directly engineered to one of the acetic acid moieties in EDTA-derived LCTs, they comprise an additional pseudoasymmetric center on the nitrogen atom that leads to two diastereomeric species in the protein-tag conjugate.⁸⁷ This leads to peak doubling observed for the tagged protein in ^1H - ^{15}N -HSQC experiments and therefore to two sets of signals and associated anisotropy parameters.⁸⁷ This issue was overcome by Leonov et al., who introduced EDTA-based LCTs with a linker moiety attached to the carbon backbone of the EDTA LCT and thus without the additional pseudoasymmetric center in the metal complex that could lead to two signal sets in recorded spectra.¹⁷⁰ By the use of DTPA-based tags, up to eight different sets of signals can be expected in theory and for two-armed LCTs, up to five different sets of signals were observed in NMR experiments (Figure 12).¹⁶⁸ In order to produce LCTs with a more rigid chelation of the lanthanide ion, enhanced PCSs and an applicability over a wide range of conditions, the development of DOTA-based LCTs for paramagnetic NMR spectroscopy was envisioned.

1.9 DOTA-based lanthanide chelating tags

DOTA-based ligands provide an ideal framework for the chelation of lanthanide ions and development of LCTs. First, DOTA-type ligands display very high binding affinities to lanthanide ions and chemical stability (equilibrium constant for the formation of $\text{Gd(III)-DOTA} = 10^{25} \text{ M}^{-1}$ at pH 7).¹⁷¹ Second, the geometry of DOTA-based LCTs allows to incorporate a linker moiety on one of the side arms and to engineer a large variety of donor atom sets.⁷⁴ Third, the incorporation of additional methyl substituents delivering a highly rigidified DOTA-

based ligand⁸¹ or the equipment of the DOTA-based ligand with two linkers⁸³ enable the detection of large PCSs.

Häussinger et al. reported in a conference abstract in the year 2005 the first use of a DOTA-based LCT for application on the dimer of the calcium dependent cell adhesion protein ECAD12 (50 kDa).^{81, 172} The LCT was based on a DOTA scaffold equipped with a linker coordinated via an amide oxygen to the metal center. Since DOTA displays an equilibrium constant for the binding of gadolinium(III) ions that is eight orders of magnitude larger than the one of EDTA, such a tag is ideally suited for characterization of proteins that require a calcium(II) chelator in the buffer, as e.g. the used ECAD12. Furthermore, the proposed LCT was suitable for application in a phosphate buffer over a wide pH range. Methyl substituents were introduced both in the cyclen scaffold and the lactic acid side arms in order to decrease the inherent motions of DOTA-derived LCTs that lead to motional averaging of PCSs (Figure 13).⁸¹

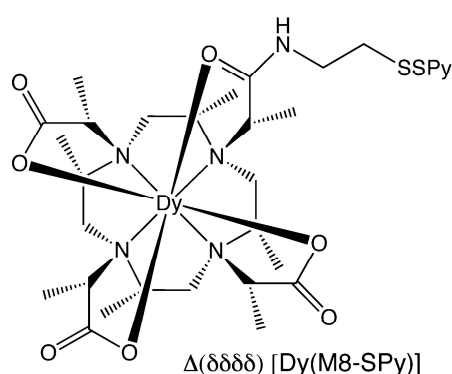


Figure 13: Structure of the rigidified DOTA-based Dy-DOTA-M8-(8S)-SSPy in TSAP conformation. Reprinted with permission from D. Häussinger, J.-R. Huang, S. Grzesiek, *J. Am. Chem. Soc.*, 2009, 131, 14761-14767. Copyright 2009 American Chemical Society.⁸¹

The methyl substituents of the cyclen scaffold adopt an equatorial-upper position¹⁷³⁻¹⁷⁴ when the DOTA-M8-(8S)-SSPy ligand is coordinated to a lanthanide ion and suppress motional averaging and therefore line broadening of NMR signals by locking the twelve-membered ring in a ($\delta\delta\delta\delta$) conformation (Figure 14).⁸¹

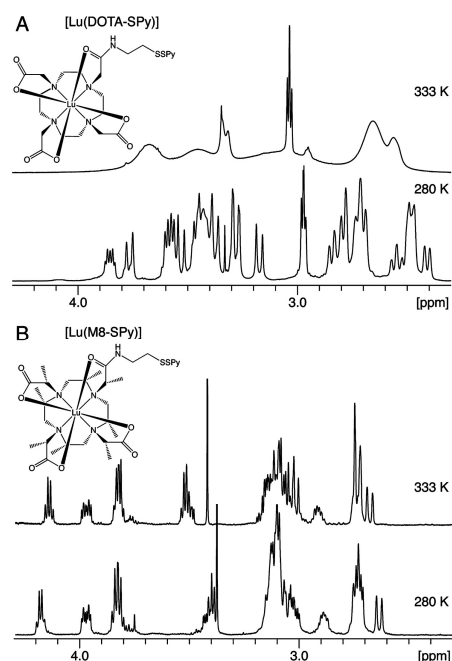


Figure 14: A) Motional averaging of NMR signals using a non-rigidified DOTA chelator. At 280 K, the internal motions are clearly decreased. B) For the methyl-substituted variant of the chelator, the motions are severely suppressed even at elevated temperatures. Reprinted with permission from D. Häussinger, J.-R. Huang, S. Grzesiek, *J. Am. Chem. Soc.*, 2009, 131, 14761-14767. Copyright 2009 American Chemical Society.⁸¹

Ln-DOTA-M8-(8S)-SSPy is a strongly paramagnetic, extremely rigidified LCT that was shown to display PCSs of over 5 ppm and RDCs over 20 Hz on an ubiquitin S57C protein construct. Besides the fact that the large magnitude of PCSs and RDCs is in an impressive range for a single-point attached LCT, the tag can, due to the high stability of the lanthanide complex, be used under extreme measurement conditions (pH = 2.5 or EDTA-containing buffer).⁸¹

Interestingly, Ln-DOTA-M8-(8S)-SSPy can show two sets of signals, which results in a complicated analysis of spectra obtained by NMR spectroscopy containing the structural restraints for protein structure determination. As shown by Opina et al., the two sets of signals have their origin in the presence of two different conformers of the LCTs, which can be interconverted to each other by rotation of the pendant arms of the chelators.¹⁷³ In the course of their investigations, the authors showed that for the Ln-DOTA-M8-(8S)-SSPy LCT both arm rotation conformers are present in significant ratios for lanthanides smaller than gadolinium and larger than thulium.¹⁷³ Therefore, the change in the ratio of the two possible conformers is dependent on the lanthanide ionic radii and therefore ultimately caused by the lanthanide contraction.¹⁷³ Furthermore, it was demonstrated that a single isolated conformer using HPLC purification partially interconverts to the other conformer and that the equilibrium is reached after approximately 4 h.¹⁷³

Since the publication of the Ln-DOTA-M8-(8S)-SSPy LCT in 2009, it was applied in a wide range of projects in our group and in collaboration with other groups. Among other projects,

the investigated examples include the localization of inhibitors on proteins,¹⁴³ the exploration of the metal position of mobile spin labels with new fitting routines¹⁷⁵ as well as studies of large protein complexes in solution.⁸⁵

Today, a large variety of DOTA-derived LCTs is available, including reduction-stable linkers, single- and two-armed LCTs as well as LCTs with linkers that are suitable for attachment to artificial amino acids.^{78-79, 83, 97, 109-110, 176-179} However, there is a large open space for improvement:

- The ligation kinetics as well as the stability of an optimal LCT have to be suitable for physiological conditions
- The distance range that is covered can be extended by larger induced anisotropy parameters
- Free copper ions in LCTs that need an azide-alkyne click reaction exclude certain protein structures

Therefore, the development of a single-armed LCT that induces large anisotropy parameters, displays fast ligation kinetics under physiological conditions, leads to reduction-stable linkages, omits the need for free lanthanide ions during protein handling and exhibits only one conformational species is highly desirable.

1.10 Formylglycine-generating enzyme

The stability of proteins in solution is crucial to the function and survival of enzymes.³³ Proteins with a stable and rigid fold in solution typically show a substantial percentage of residues in secondary structural elements, as e.g. alpha-helices or beta-sheets. In order to fulfil the important catalytic and regulatory functions of enzymes, the specifically defined structure of the active site is the essential feature for stabilization of the transition state and hence catalytic activity.¹⁸⁰⁻¹⁸¹ Furthermore, internal dynamics are more and more recognized as a driving force for catalysis in enzymes.¹⁸²⁻¹⁸³

FGE exhibits a special fold in the protein world and displays several interesting structural properties. Common catalytically active copper proteins as e.g. particulate methane monooxygenase (pMMO),¹⁸⁴ Lytic polysaccharide monooxygenases (LPMOs),¹⁸⁵ peptidylglycine- α -hydroxylating monooxygenase (PHM)¹⁸⁶, dopamine- β -hydroxylase (DBH)¹⁸⁷ and copper amine oxidase (CAO)¹⁸⁸ display N-containing ligand dominated coordination spheres. In contrast, copper trafficking proteins as e.g. CueR¹⁸⁹ and HAH1 metallochaperone protein¹⁹⁰ contain a bislinear cysteine coordination motif. FGE is to date the only observed highly catalytically active protein bearing two cysteines coordinating the copper cofactor in a

bilinear geometry. Furthermore, based on the solved crystal structure, it is known that FGE contains only few extended regions of the classical secondary structure elements such as alpha-helices and beta-sheets.¹⁹¹ In contrast to the intuitive assumption that this should lead to a decreased stability of the enzyme, FGE shows a very high temperature stability of 80 °C. Since FGE contains an unusual high number of prolines (9%, average in Swiss Prot. data bank: 4.7%), the high melting point potentially originates in the relative lower entropy of the unfolded structure vs. the folded protein when compared to other proteins.¹⁹² FGE catalyzes the oxygen-dependent conversion of specific cysteine residues of a protein to formylglycine using copper as cofactor.^{191, 193-194} More specifically, FGE is responsible for activation of type I sulfatases, a class of enzymes that relies on a formylglycine residue for the hydrolysis of sulfate ester bonds (Figure 15).¹⁹⁵

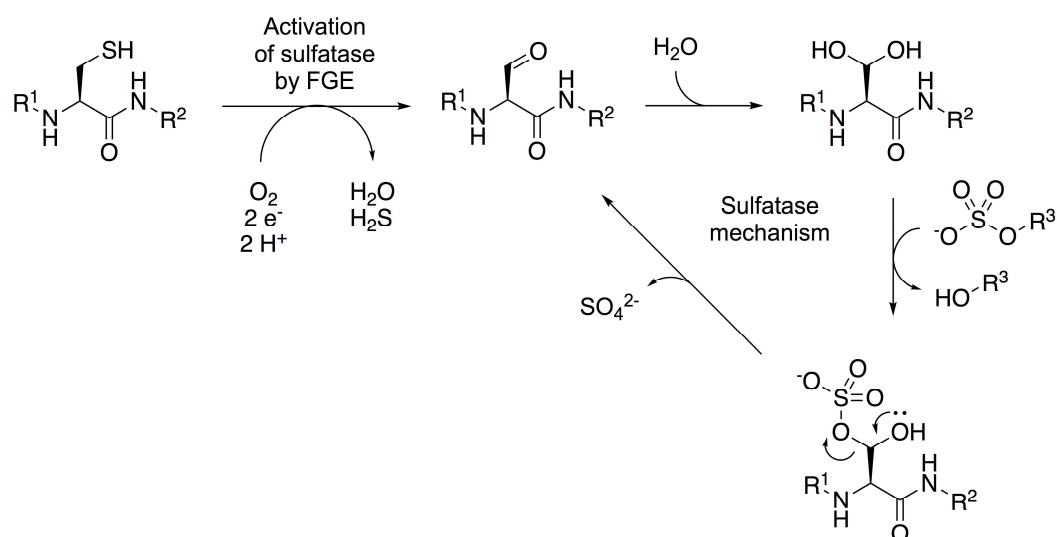


Figure 15: FGE-dependent activation and the mechanism of the following catalytic activity of a type I sulfatase.¹⁹⁵

Upon generating the aldehyde functionality in the sulfatase by FGE, the binding of a water molecule yields a geminal diol functionality that can be used by the sulfatase to cleave sulfate ester bonds of various substrates, as e.g. steroids, proteins and carbohydrates. After the successful cleavage of the sulfate ester bond, the sulfate ion, which was transferred from the substrate to the enzyme is released and thereby the aldehyde functionality of the sidechain of the catalytically active residue in the sulfatase is restored.

Interestingly, a lack of FGE present in humans is linked to the severe disease MSD, since the absence of FGE prohibits the activation of type I sulfatases by absence of the conversion of the crucial cysteine residue in the sulfatase to a formylglycine sidechain.¹⁹⁶⁻¹⁹⁹ The identification of disease-related mutations in human patients, which affect the stability and folding as well as the catalytic activity of the enzyme, provides an intriguing opportunity to elucidate the molecular basis of the multiple sulfatase deficiency by acquisition of 2D ¹H-¹⁵N

HSQC experiments in combination with a backbone assignment of the protein. Since there is a high interest in a characterization of the FGE mechanism as well as of native and disease-related variants of the FGE protein, an NMR study was envisioned in this PhD project.

FGE incorporates an amount of 23.3% charged amino acids (average in Uniprot database: 25.8%) and exhibits 6 highly conserved salt bridges with an average bridged distance of 2.9 Å.¹⁹¹ FGE adopts a compact, globular fold, stabilized by H-bonding networks, salt bridges, hydrophobic interactions and three metal ions (Figure 16).

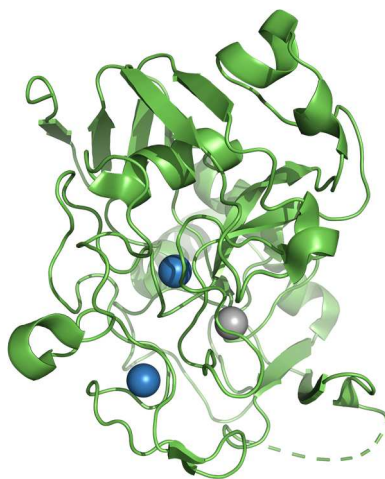


Figure 16: Globular fold of the FGE protein in complex with a Ag(I) ion (grey) in its crystal structure (PDB 5NXL).¹⁹¹

Despite the lack of extensive amounts of secondary structure in its core and around the active site, the FGE protein adopts a well-defined and folded tertiary structure and was crystallized for the holoenzyme containing Ag(I) or Cd(II).¹⁹¹ Further insights were provided by the structure of FGE in complex with copper reported by Appel et al.²⁰⁰ The interesting structural and catalytic properties of the FGE protein render it a highly promising target for NMR investigations of both its protein fold and mechanism.

1.11 Malondialdehyde Adducts of Nucleosides

T lymphocytes can destroy diseased cells or mutated cancer cells within patients.²⁰¹ The T lymphocytes recognize antigens that are complexed and presented by proteins on the surface of a cell.²⁰² In cancer cells, the accelerated metabolism leads to oxidative stress within the cells.²⁰³ The formed adducts of molecules with reactive oxygen species during oxidative stress could serve as antigens for the recognition of cancer cells by the immune system. While it was shown for riboflavin-like molecules that these antigens in complex with the MHC-I-related molecule MR1 can be recognized by T lymphocytes on the surface of the cell,²⁰² the intriguing

question remains if nucleic acid adducts caused by oxidative damages can be characterized and potentially used as a vaccine against cancer. Due to the oxidative stress and enhanced metabolism in cancer cells, the concentration of malondialdehyde (MDA) increases in the diseased cells. The toxic and strongly reactive MDA molecules lead then to modifications and damages on other substrates, e.g. proteins and nucleic acids.²⁰⁴⁻²⁰⁵ Besides the cyclic MDA adduct of deoxyguanosine (M1dG), a large variety of other oxidative damages can be expected, e.g. the modification of adenosine, cytidine and guanine.^{204, 206}

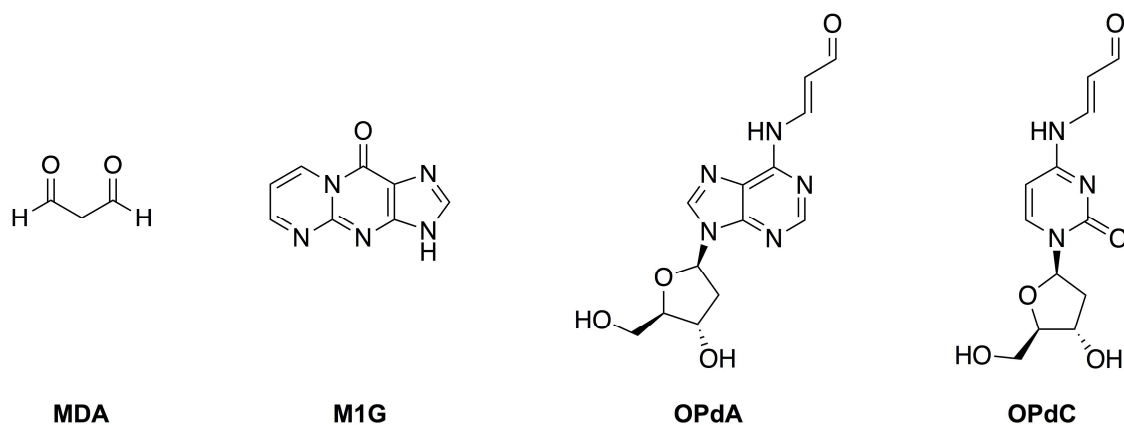


Figure 17: Structure of MDA and its adducts with guanine (M1G), deoxyadenosine (OPdA) and deoxycytidine (OPdC).

The synthesis of MDA adducts of DNA bases and the subsequent investigation of their biological properties can provide further insights in this field (Figure 17). Furthermore, although this dream remains far in the future, the recognition of specific oxidative damages in cancer cells by the human immune system could lead to a vaccine against cancer by training of the immune system with cancer-derived antigens. Exploitation of the endogenous immune mechanisms of the human body to get rid of cancer cells would constitute a highly elegant way to deal with the increasingly old society and the associated rise in cancer prevalence.

2 Research goal

The first part of this thesis deals with the development of significantly improved generations of LCTs, since an LCT comprising all previously mentioned favorable properties is of high need (high affinity for lanthanide(III) ions, reduction-stable linker, fast ligation under diluted conditions, one conformational species and large induced anisotropic effects). The aim was to start with the restriction to only one conformational species and subsequent equipment of this LCT with isopropyl substituents in order to reduce the translational freedom of the LCT. Subsequently, we scheduled to design an activator moiety that provides a reduction-stable linkage upon ligation of the LCT to the protein. Last but not least, the further restriction of the rotational degrees of freedom of an LCT with only one conformational species, reduction-stable linker and suppressed translational averaging of PCSs was envisioned in order to strongly boost the anisotropic effects obtained in protein NMR experiments.

Due to the fast developments in the field of LCTs, an up-to-date review comprising all different aspects of this research field was highly sought-after. Therefore, an extensive review about the design and applications of LCTs for the investigation of proteins, nucleic acids, carbohydrate and complexes thereof was planned for publication in *Progress in Nuclear Magnetic Resonance Spectroscopy*. Furthermore, a book chapter about the application of LCTs for the localization of ligands was scheduled in order to contribute to *Comprehensive Coordination Chemistry III*.

In the third part of this thesis, the transfer of the LCT methodology from the protein to the RNA world is covered. Currently, there is no approach available for the covalent and direct conjugation of LCTs to RNA targets in order to exploit paramagnetic structural restraints for elucidation of structure, dynamics and ligand binding of RNA strands. The development of a novel and general applicable LCT that enables to record paramagnetic structural restraints on RNA molecules by analysis of the obtained PCS and RDC data is thus of urgent need and would provide the opportunity to significantly boost the investigation of the rich structural diversity of RNA molecules as well as RNA-RNA and RNA-ligand complexes. Accordingly, the design, synthesis and performance evaluation of such a chelator was included into the research plan for this thesis.

Fourth, the elucidation of structural properties, dynamics and mechanism of the FGE protein was envisioned. Given the unique fold and active site architecture of this protein in combination with the ability of protein NMR spectroscopy to provide residue-specific insights into the solution state structural properties and dynamics of the enzyme, a thorough study was

scheduled. In order to gain more knowledge about this special enzyme, the catalytically involved residues and their function in the active site and the stability as well as flexibility of the protein in different metalation states were planned to be investigated. Furthermore, disease-related mutants were targeted in order to provide a molecular basis for the occurrence of specific phenotypes in patients with multiple sulfatase deficiency.

For the fifth part of this thesis, it was planned to synthesize and characterize four cancer-related, nucleic acid-derived antigens. Based on the initial biological activity of an unknown MDA-, and thus oxidative damage-related adduct, the structure and the synthesis of both the identified antigen and three synthetic antigens were scheduled. The binding of these antigens to the MHC-I-related molecule MR1 and potential subsequent activation of MR1T lymphocytes were targeted for investigation. We speculated that the elucidation of the identity of these adducts could lead to more detailed insights about oxidative damages of nucleic acid-derived molecules and later to the development of a potential vaccine against cancer.

References

1. H. L. Anderson, C. A. Hunter, M. N. Meah, J. K. M. Sanders, *J. Am. Chem. Soc.*, **1990**, *112*, 5780-5789.
2. G. Bringmann, M. Heubes, M. Breuning, L. Göbel, M. Ochse, B. Schöner, O. Schupp, *J. Org. Chem.*, **2000**, *65*, 722-728.
3. J. Rotzler, H. Gsellinger, M. Neuburger, D. Vonlanthen, D. Häussinger, M. Mayor, *Org. Biomol. Chem.*, **2011**, *9*, 86-91.
4. J. Rotzler, H. Gsellinger, A. Bihlmeier, M. Gantenbein, D. Vonlanthen, D. Häussinger, W. Klopfer, M. Mayor, *Org. Biomol. Chem.*, **2013**, *11*, 110-118.
5. D. Eliezer, *Curr. Opin. Struct. Biol.*, **2009**, *19*, 23-30.
6. S. B. Shuker, P. J. Hajduk, R. P. Meadows, S. W. Fesik, *Science*, **1996**, *274*, 1531-1534.
7. B. Meyer, T. Peters, *Angew. Chem. Int. Ed.*, **2003**, *42*, 864-890.
8. J. L. Markley, W. M. Westler, *Arch. Biochem. Biophys.*, **2017**, *628*, 3-16.
9. R. Hurwitz, *J. Magn. Reson. Imag.*, **1992**, *2*, 131-133.
10. G. Breit, I. I. Rabi, *Phys. Rev.*, **1931**, *38*, 2082-2083.
11. I. I. Rabi, J. R. Zacharias, S. Millman, P. Kusch, *Phys. Rev.*, **1938**, *53*, 318.
12. M. A. Shampo, R. A. Kyle, D. P. Steensma, *Mayo Clin. Proc.*, **2012**, *87*, e11.
13. R. K. Wangsness, F. Bloch, *Phys. Rev.*, **1953**, *89*, 728-739.
14. F. Bloch, W. W. Hansen, M. Packard, *Phys. Rev.*, **1946**, *70*, 474-485.
15. F. Bloch, *Phys. Rev.*, **1946**, *70*, 460-474.
16. S. R. Hartmann, E. L. Hahn, *Phys. Rev.*, **1962**, *128*, 2042-2053.
17. M. Emshwiller, E. L. Hahn, D. Kaplan, *Phys. Rev.*, **1960**, *118*, 414-424.
18. M. Bloom, E. L. Hahn, B. Herzog, *Phys. Rev.*, **1955**, *97*, 1699-1709.
19. E. L. Hahn, D. E. Maxwell, *Phys. Rev.*, **1952**, *88*, 1070-1084.
20. E. L. Hahn, *Phys. Rev.*, **1950**, *77*, 297-298.
21. P. Mansfield, *Angew. Chem. Int. Ed.*, **2004**, *43*, 5456-5464.
22. P. C. Lauterbur, *Nature*, **1973**, *242*, 190-191.
23. M. J. Stehling, A. M. Howseman, R. J. Ordidge, B. Chapman, R. Turner, R. Coxon, P. Glover, P. Mansfield, R. E. Coupland, *Radiology*, **1989**, *170*, 257-263.
24. L. Braunschweiler, R. R. Ernst, *J. Magn. Reson.*, **1983**, *53*, 521-528.
25. W. P. Aue, E. Bartholdi, R. R. Ernst, *J. Chem. Phys.*, **1976**, *64*, 2229-2246.
26. A. Kumar, R. R. Ernst, K. Wüthrich, *Biochem. Biophys. Res. Commun.*, **1980**, *95*, 1-6.
27. C. Griesinger, G. Otting, K. Wuethrich, R. R. Ernst, *J. Am. Chem. Soc.*, **1988**, *110*, 7870-7872.
28. S. Macura, Y. Huang, D. Suter, R. R. Ernst, *J. Magn. Reson.*, **1981**, *43*, 259-281.
29. A. Bax, R. Freeman, *J. Magn. Reson.*, **1981**, *44*, 542-561.
30. G. A. Morris, R. Freeman, *J. Magn. Reson.*, **1978**, *29*, 433-462.
31. M. H. Levitt, R. Freeman, *J. Magn. Reson.*, **1981**, *43*, 502-507.
32. A. Bax, R. Freeman, G. Morris, *J. Magn. Reson.*, **1981**, *42*, 164-168.
33. B. Alberts, Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P., *Molecular Biology of the Cell, 4th edition*. Garland Science: New York, 2002.
34. K. Wüthrich, *Angew. Chem. Int. Ed.*, **2003**, *42*, 3340-3363.
35. S. Sugio, A. Kashima, S. Mochizuki, M. Noda, K. Kobayashi, *Protein Eng.*, **1999**, *12*, 439-446.
36. K. Inomata, A. Ohno, H. Tochio, S. Isogai, T. Tenno, I. Nakase, T. Takeuchi, S. Futaki, Y. Ito, H. Hiroaki, M. Shirakawa, *Nature*, **2009**, *458*, 106-109.
37. D. Sakakibara, A. Sasaki, T. Ikeya, J. Hamatsu, T. Hanashima, M. Mishima, M. Yoshimasu, N. Hayashi, T. Mikawa, M. Wälchli, B. O. Smith, M. Shirakawa, P. Güntert, Y. Ito, *Nature*, **2009**, *458*, 102.
38. P. Selenko, Z. Serber, B. Gadea, J. Ruderman, G. Wagner, *Proc. Natl. Acad. Sci.*, **2006**, *103*, 11904-11909.
39. Z. Serber, V. Dötsch, *Biochemistry*, **2001**, *40*, 14317-14323.
40. K. Wüthrich, *Nat. Struct. Biol.*, **2001**, *8*, 923-925.
41. M. P. Williamson, T. F. Havel, K. Wüthrich, *J. Mol. Biol.*, **1985**, *182*, 295-315.
42. M. Piotto, V. Saudek, V. Sklenář, *J. Biomol. NMR*, **1992**, *2*, 661-665.
43. T. L. Hwang, A. J. Shaka, *J. Magn. Reson., Series A*, **1995**, *112*, 275-279.
44. G. A. Morris, *J. Am. Chem. Soc.*, **1980**, *102*, 428-429.
45. L. E. Kay, M. Ikura, R. Tschudin, A. Bax, *J. Magn. Reson.*, **1990**, *89*, 496-514.
46. M. Ikura, L. E. Kay, A. Bax, *Biochemistry*, **1990**, *29*, 4659-4667.
47. L. E. Kay, G. M. Clore, A. Bax, A. M. Gronenborn, *Science*, **1990**, *249*, 411.
48. M. Sattler, S. W. Fesik, *Structure*, **1996**, *4*, 1245-1249.

49. K. Pervushin, R. Riek, G. Wider, K. Wüthrich, *Proc. Natl. Acad. Sci.*, **1997**, *94*, 12366-12371.
50. M. Salzmann, G. Wider, K. Pervushin, H. Senn, K. Wüthrich, *J. Am. Chem. Soc.*, **1999**, *121*, 844-848.
51. M. Salzmann, G. Wider, K. Pervushin, K. Wuthrich, *J. Biomol. NMR*, **1999**, *15*, 181-184.
52. M. Salzmann, K. Pervushin, G. Wider, H. Senn, K. Wuthrich, *Proc. Natl. Acad. Sci.*, **1998**, *95*, 13585-13590.
53. P. Neudecker, P. Lundström, L. E. Kay, *Biophys. J.*, **2009**, *96*, 2045-2054.
54. P. Schanda, Ě. Kupče, B. Brutscher, *J. Biomol. NMR*, **2005**, *33*, 199-211.
55. D. Rovnyak, D. P. Frueh, M. Sastry, Z.-Y. J. Sun, A. S. Stern, J. C. Hoch, G. Wagner, *J. Magn. Reson.*, **2004**, *170*, 15-21.
56. P. Rößler, D. Mathieu, A. D. Gossert, *Angew. Chem. Int. Ed.*, **2020**, *59*, 19329-19337.
57. R. Rosenzweig, N. B. Nillegoda, M. P. Mayer, B. Bukau, *Nat. Rev. Mol. Cell Biol.*, **2019**, *20*, 665-680.
58. S. Ambadipudi, J. Biernat, D. Riedel, E. Mandelkow, M. Zweckstetter, *Nat. Comm.*, **2017**, *8*, 275.
59. T. Madl, M. Sattler NMR Methodologies for the Analysis of Protein-Protein Interactions. In *NMR of Biomolecules*, Wiley: 2012; p 173-194.
60. Z. Yu, P. Li, K. M. Merz, *Biochemistry*, **2017**, *56*, 2349-2362.
61. G. M. Clore, A. M. Gronenborn, *J. Magn. Reson.*, **1982**, *48*, 402-417.
62. T. Carlomagno, *J. Magn. Reson.*, **2014**, *241*, 126-136.
63. P. H. J. Keizers, M. Ubbink, *Prog. Nucl. Magn. Reson. Spectrosc.*, **2011**, *58*, 88-96.
64. J. R. Bothe, E. N. Nikolova, C. D. Eichhorn, J. Chugh, A. L. Hansen, H. M. Al-Hashimi, *Nat. Methods*, **2011**, *8*, 919-931.
65. J. Rinnenthal, J. Buck, J. Ferner, A. Wacker, B. Fürtig, H. Schwalbe, *Acc. Chem. Res.*, **2011**, *44*, 1292-1301.
66. J. Flinders, T. Dieckmann, *Prog. Nucl. Magn. Reson. Spectrosc.*, **2006**, *48*, 137-159.
67. C. Dominguez, M. Schubert, O. Duss, S. Ravindranathan, F. H. T. Allain, *Prog. Nucl. Magn. Reson. Spectrosc.*, **2011**, *58*, 1-61.
68. B. Fürtig, C. Richter, J. Wöhnert, H. Schwalbe, *ChemBioChem*, **2003**, *4*, 936-962.
69. M. Nitzan, Karen M. Wassarman, O. Biham, H. Margalit, *Biophys. J.*, **2014**, *106*, 1205-1214.
70. K. I. Kristiansen, R. Weel-Sneve, J. A. Booth, M. Bjørås, *RNA*, **2016**, *22*, 1739-1749.
71. K. D. Warner, C. E. Hajdin, K. M. Weeks, *Nat. Rev. Drug Discov.*, **2018**, *17*, 547-558.
72. J. Kremser, E. Strebiter, R. Plangger, M. A. Juen, F. Nußbaumer, H. Glasner, K. Breuker, C. Kreutz, *Chem. Commun.*, **2017**, *53*, 12938-12941.
73. S. A. McCallum, A. Pardi, *J. Mol. Biol.*, **2003**, *326*, 1037-1050.
74. C. Nitsche, G. Otting, *Prog. Nucl. Magn. Reson. Spectrosc.*, **2017**, *98-99*, 20-49.
75. W.-M. Liu, M. Overhand, M. Ubbink, *Coord. Chem. Rev.*, **2014**, *273-274*, 2-12.
76. B.-B. Pan, F. Yang, Y. Ye, Q. Wu, C. Li, T. Huber, X.-C. Su, *Chem. Commun.*, **2016**, *52*, 10237-10240.
77. T. Müntener, D. Häussinger, P. Selenko, F.-X. Theillet, *J. Phys. Chem. Lett.*, **2016**, *7*, 2821-2825.
78. Y. Hikone, G. Hirai, M. Mishima, K. Inomata, T. Ikeya, S. Arai, M. Shirakawa, M. Sodeoka, Y. Ito, *J. Biomol. NMR*, **2016**, *66*, 99-110.
79. C. T. Loh, K. Ozawa, K. L. Tuck, N. Barlow, T. Huber, G. Otting, B. Graham, *Bioconjug. Chem.*, **2013**, *24*, 260-268.
80. X.-C. Su, H. Liang, K. V. Loscha, G. Otting, *J. Am. Chem. Soc.*, **2009**, *131*, 10352-10353.
81. D. Häussinger, J.-R. Huang, S. Grzesiek, *J. Am. Chem. Soc.*, **2009**, *131*, 14761-14767.
82. X.-C. Su, B. Man, S. Beeren, H. Liang, S. Simonsen, C. Schmitz, T. Huber, B. A. Messerle, G. Otting, *J. Am. Chem. Soc.*, **2008**, *130*, 10486-10487.
83. P. H. J. Keizers, A. Saragliadis, Y. Hiruma, M. Overhand, M. Ubbink, *J. Am. Chem. Soc.*, **2008**, *130*, 14802-14812.
84. G. Pintacuda, A. Y. Park, M. A. Keniry, N. E. Dixon, G. Otting, *J. Am. Chem. Soc.*, **2006**, *128*, 3696-3702.
85. K. D. Brewer, T. Bacaj, A. Cavalli, C. Camilloni, J. D. Swarbrick, J. Liu, A. Zhou, P. Zhou, N. Barlow, J. Xu, A. B. Seven, E. A. Prinslow, R. Voleti, D. Häussinger, A. M. J. J. Bonvin, D. R. Tomchick, M. Vendruscolo, B. Graham, T. C. Südhof, J. Rizo, *Nat. Struct. Mol. Biol.*, **2015**, *22*, 555-564.
86. V. Gaponenko, A. S. Altieri, J. Li, R. A. Byrd, *J. Biomol. NMR*, **2002**, *24*, 143-148.
87. T. Ikegami, L. Verdier, P. Sakhaii, S. Grimme, B. Pescatore, K. Saxena, K. M. Fiebig, C. Griesinger, *J. Biomol. NMR*, **2004**, *29*, 339-349.
88. F. Peters, M. Maestre-Martinez, A. Leonov, L. Kovacic, S. Becker, R. Boelens, C. Griesinger, *J. Biomol. NMR*, **2011**, *51*, 329-337.
89. J. Wöhnert, K. J. Franz, M. Nitz, B. Imperiali, H. Schwalbe, *J. Am. Chem. Soc.*, **2003**, *125*, 13338-13339.
90. I. Bertini, P. Turano, A. J. Vila, *Chem. Rev.*, **1993**, *93*, 2833-2932.

91. G. M. Clore, J. Iwahara, *Chem. Rev.*, **2009**, *109*, 4108-4139.
92. J. Iwahara, G. M. Clore, *Nature*, **2006**, *440*, 1227-1230.
93. I. Bertini, C. Luchinat, G. Parigi, R. Pierattelli, *ChemBioChem*, **2005**, *6*, 1536-1549.
94. G. Pintacuda, M. John, X.-C. Su, G. Otting, *Acc. Chem. Res.*, **2007**, *40*, 206-212.
95. L. Banci, I. Bertini, K. L. Bren, M. A. Cremonini, H. B. Gray, C. Luchinat, P. Turano, *J. Biol. Inorg. Chem.*, **1996**, *1*, 117-126.
96. G. Otting, *Annu. Rev. Biophys.*, **2010**, *39*, 387-405.
97. M. D. Lee, C. T. Loh, J. Shin, S. Chhabra, M. L. Dennis, G. Otting, J. D. Swarbrick, B. Graham, *Chem. Sci.*, **2015**, *6*, 2614-2624.
98. B. J. G. Pearce, S. Jabar, C.-T. Loh, M. Szabo, B. Graham, G. Otting, *J. Biomol. NMR*, **2017**, *68*, 19-32.
99. M. Strickland, C. D. Schwieters, C. Göbl, A. C. L. Opina, M.-P. Strub, R. E. Swenson, O. Vasalatiy, N. Tjandra, *J. Biomol. NMR*, **2016**, *66*, 125-139.
100. K. Tu, M. Gochin, *J. Am. Chem. Soc.*, **1999**, *121*, 9276-9285.
101. B. Liang, J. H. Bushweller, L. K. Tamm, *J. Am. Chem. Soc.*, **2006**, *128*, 4389-4397.
102. J. R. Gillespie, D. Shortle, *J. Mol. Biol.*, **1997**, *268*, 158-169.
103. J. R. Gillespie, D. Shortle, *J. Mol. Biol.*, **1997**, *268*, 170-184.
104. J. Iwahara, C. D. Schwieters, G. M. Clore, *J. Am. Chem. Soc.*, **2004**, *126*, 5879-5896.
105. J. Iwahara, C. Tang, G. Marius Clore, *J. Magn. Reson.*, **2007**, *184*, 185-195.
106. L. W. Donaldson, N. R. Skrynnikov, W.-Y. Choy, D. R. Muhandiram, B. Sarkar, J. D. Forman-Kay, L. E. Kay, *J. Am. Chem. Soc.*, **2001**, *123*, 9843-9847.
107. C. A. Barnes, Y. Shen, J. Ying, Y. Takagi, D. A. Torchia, J. R. Sellers, A. Bax, *J. Am. Chem. Soc.*, **2019**, *141*, 9004-9017.
108. S. C. Chiliveri, J. M. Louis, R. Ghirlando, J. L. Baber, A. Bax, *J. Am. Chem. Soc.*, **2018**, *140*, 34-37.
109. M. D. Lee, M. L. Dennis, J. D. Swarbrick, B. Graham, *Chem. Commun.*, **2016**, *52*, 7954-7957.
110. M. D. Lee, M. L. Dennis, B. Graham, J. D. Swarbrick, *Chem. Commun.*, **2017**, *53*, 13205-13208.
111. K. Chen, N. Tjandra, *Top. Curr. Chem.*, **2012**, *326*, 47-67.
112. R. Barbieri, I. Bertini, G. Cavallaro, Y.-M. Lee, C. Luchinat, A. Rosato, *J. Am. Chem. Soc.*, **2002**, *124*, 5581-5587.
113. T. Saio, K. Ishimori, *Biochim. Biophys. Acta. Gen. Subj.*, **2020**, *1864*, 129332.
114. B. Bleaney, C. M. Dobson, B. A. Levine, R. B. Martin, R. J. P. Williams, A. V. Xavier, *J. Chem. Soc., Chem. Commun.*, **1972**, 791b-793.
115. B. Bleaney, *J. Magn. Reson.*, **1972**, *8*, 91-100.
116. A. M. Funk, K.-L. N. A. Finney, P. Harvey, A. M. Kenwright, E. R. Neil, N. J. Rogers, P. Kanthi Senanayake, D. Parker, *Chem. Sci.*, **2015**, *6*, 1655-1662.
117. E. A. Sutturina, K. Mason, C. F. G. C. Gerales, I. Kuprov, D. Parker, *Angew. Chem. Int. Ed.*, **2017**, *56*, 12215-12218.
118. P. V. Demarco, T. K. Elzey, R. B. Lewis, E. Wenkert, *J. Am. Chem. Soc.*, **1970**, *92*, 5737-5739.
119. A. F. Cockerill, D. M. Rackham, *Tetrahedron Lett.*, **1970**, *11*, 5149-5152.
120. G. M. Whitesides, D. W. Lewis, *J. Am. Chem. Soc.*, **1970**, *92*, 6979-6980.
121. H. L. Goering, J. N. Eikenberry, G. S. Koermer, *J. Am. Chem. Soc.*, **1971**, *93*, 5913-5914.
122. C. Nitsche, G. Otting, *Prog. Nucl. Magn. Reson. Spectrosc.*, **2017**, *98-99*, 20-49.
123. M. Gochin, H. Roder, *Protein Sci.*, **1995**, *4*, 296-305.
124. Z. Gong, C. D. Schwieters, C. Tang, *Methods*, **2018**, *148*, 48-56.
125. J. Gao, E. Liang, R. Ma, F. Li, Y. Liu, J. Liu, L. Jiang, C. Li, H. Dai, J. Wu, X. Su, W. He, K. Ruan, *Angew. Chem. Int. Ed.*, **2017**, *56*, 12982-12986.
126. M. John, G. Pintacuda, A. Y. Park, N. E. Dixon, G. Otting, *J. Am. Chem. Soc.*, **2006**, *128*, 12910-12916.
127. J.-Y. Guan, P. H. J. Keizers, W.-M. Liu, F. Löhr, S. P. Skinner, E. A. Heeneman, H. Schwalbe, M. Ubbink, G. Siegal, *J. Am. Chem. Soc.*, **2013**, *135*, 5859-5868.
128. T. Saio, K. Ogura, K. Shimizu, M. Yokochi, T. R. Burke, F. Inagaki, *J. Biomol. NMR*, **2011**, *51*, 395-408.
129. U. Brath, S. I. Swamy, A. X. Veiga, C.-C. Tung, F. Van Petegem, M. Erdélyi, *J. Am. Chem. Soc.*, **2015**, *137*, 11391-11398.
130. C. Öster, S. Kosol, C. Hartmüller, J. M. Lamley, D. Iuga, A. Oss, M.-L. Org, K. Vanatalu, A. Samoson, T. Madl, J. R. Lewandowski, *J. Am. Chem. Soc.*, **2017**, *139*, 12165-12174.
131. C. Göbl, M. Resch, M. Strickland, C. Hartmüller, M. Viertler, N. Tjandra, T. Madl, *Angew. Chem. Int. Ed.*, **2016**, *55*, 14847-14851.
132. M. Vonci, K. Mason, E. A. Sutturina, A. T. Frawley, S. G. Worswick, I. Kuprov, D. Parker, E. J. L. McInnes, N. F. Chilton, *J. Am. Chem. Soc.*, **2017**, *139*, 14166-14172.

133. A. Roca-Sabio, M. Regueiro-Figueroa, D. Esteban-Gómez, A. de Blas, T. Rodríguez-Blas, C. Platas-Iglesias, *Comput. Theor. Chem.*, **2012**, *999*, 93-104.
134. G. Hong, M. Dolg, L. Li, *Chem. Phys. Lett.*, **2001**, *334*, 396-402.
135. M. Purgel, Z. Baranyai, A. de Blas, T. Rodríguez-Blas, I. Bányai, C. Platas-Iglesias, I. Tóth, *Inorg. Chem.*, **2010**, *49*, 4370-4382.
136. G. Schreckenbach, T. Ziegler, *J. Phys. Chem. A*, **1997**, *101*, 3388-3399.
137. S. Per, P. E. A. Turchi, A. Landa, V. Lordi, *J. Phys. Condens. Matter*, **2014**, *26*, 416001.
138. S. Grimmel, G. Schoendorff, A. K. Wilson, *J. Chem. Theory Comput.*, **2016**, *12*, 1259-1266.
139. L. E. Aebersold, S. H. Yuwono, G. Schoendorff, A. K. Wilson, *J. Chem. Theory Comput.*, **2017**, *13*, 2831-2839.
140. I. Bertini, C. Luchinat, G. Parigi, *Prog. Nucl. Magn. Reson. Spectrosc.*, **2002**, *40*, 249-273.
141. C. Schmitz, M. J. Stanton-Cook, X. C. Su, G. Otting, T. Huber, *J. Biomol. NMR*, **2008**, *41*, 179-189.
142. K. B. Pilla, G. Otting, T. Huber, *J. Mol. Biol.*, **2016**, *428*, 522-532.
143. K. Zimmermann, D. Joss, T. Müntener, E. S. Nogueira, M. Schäfer, L. Knörr, F. W. Monnard, D. Häussinger, *Chem. Sci.*, **2019**, *10*, 5064-5072.
144. J. H. Prestegard, H. M. Al-Hashimi, J. R. Tolman, *Q. Rev. Biophys.*, **2000**, *33*, 371-424.
145. J. R. Tolman, K. Ruan, *Chem. Rev.*, **2006**, *106*, 1720-1736.
146. C. A. Bewley, G. M. Clore, *J. Am. Chem. Soc.*, **2000**, *122*, 6009-6016.
147. F. Delhommel, F. Gabel, M. Sattler, *J. Mol. Biol.*, **2020**, *432*, 2890-2912.
148. X. Wang, J. P. Kirkpatrick, H. M. M. Launay, A. de Simone, D. Häussinger, C. M. Dobson, M. Vendruscolo, L. D. Cabrita, C. A. Waudby, J. Christodoulou, *Sci. Rep.*, **2019**, *9*, 13528.
149. D. MacDonald, P. Lu, *Curr. Opin. Struct. Biol.*, **2002**, *12*, 337-343.
150. G.-W. Li, H. Liu, F. Qiu, X.-J. Wang, X.-X. Lei, *Nat. Prod. Bioprospect.*, **2018**, *8*, 279-295.
151. G. J. Stasiuk, N. J. Long, *Chem. Commun.*, **2013**, *49*, 2732-2746.
152. H. R. Maেকে, J. C. Reubi, *J. Nucl. Med.*, **2011**, *52*, 841-844.
153. P. Verwilst, S. Park, B. Yoon, J. S. Kim, *Chem. Soc. Rev.*, **2015**, *44*, 1791-1806.
154. E. F. Ullman, J. H. Osiecki, D. G. B. Boocock, R. Darcy, *J. Am. Chem. Soc.*, **1972**, *94*, 7049-7059.
155. M. D. Rabenstein, Y. K. Shin, *Proc. Natl. Acad. Sci.*, **1995**, *92*, 8239-8243.
156. J. L. Battiste, G. Wagner, *Biochemistry*, **2000**, *39*, 5355-5365.
157. L. Helm, *Prog. Nucl. Magn. Reson. Spectrosc.*, **2006**, *49*, 45-64.
158. J. Blahut, P. Hermann, Z. Tosner, C. Platas-Iglesias, *Phys. Chem. Chem. Phys.*, **2017**, *19*, 26662-26671.
159. C. D. Barry, A. C. T. North, J. A. Glasel, R. J. P. Williams, A. V. Xavier, *Nature*, **1971**, *232*, 236-245.
160. I. Bertini, A. Donaire, B. Jimenez, C. Luchinat, G. Parigi, M. Piccioli, L. Poggi, *J. Biomol. NMR*, **2001**, *21*, 85-98.
161. T. Saio, K. Ogura, M. Yokochi, Y. Kobashigawa, F. Inagaki, *J. Biomol. NMR*, **2009**, *44*, 157-166.
162. M. Nitz, M. Sherawat, K. J. Franz, E. Peisach, K. N. Allen, B. Imperiali, *Angew. Chem. Int. Ed.*, **2004**, *43*, 3682-3685.
163. M. Nitz, K. J. Franz, R. L. Maglathlin, B. Imperiali, *ChemBioChem*, **2003**, *4*, 272-276.
164. K. J. Franz, M. Nitz, B. Imperiali, *ChemBioChem*, **2003**, *4*, 265-271.
165. C. Ma, S. J. Opella, *J. Magn. Reson.*, **2000**, *146*, 381-384.
166. X.-C. Su, K. McAndrew, T. Huber, G. Otting, *J. Am. Chem. Soc.*, **2008**, *130*, 1681-1687.
167. X.-C. Su, T. Huber, N. E. Dixon, G. Otting, *ChemBioChem*, **2006**, *7*, 1599-1604.
168. M. Prudêncio, J. Rohovec, J. A. Peters, E. Tocheva, M. J. Boulanger, M. E. P. Murphy, H.-J. Hupkes, W. Kisters, A. Impagliazzo, M. Ubbink, *Chem. Eur. J.*, **2004**, *10*, 3252-3260.
169. G. Pintacuda, A. Moshref, A. Leonchiks, A. Sharipo, G. Otting, *J. Biomol. NMR*, **2004**, *29*, 351-361.
170. A. Leonov, B. Voigt, F. Rodriguez-Castañeda, P. Sakhaii, C. Griesinger, *Chem. Eur. J.*, **2005**, *11*, 3342-3348.
171. M. F. Tweedle, J. J. Hagan, K. Kumar, S. Mantha, C. A. Chang, *Magn. Reson. Imaging*, **1991**, *9*, 409-415.
172. D. Häussinger, S. Grzesiek, *Conference Abstract Book*, **2005**, *1*, 1.
173. A. C. L. Opina, M. Strickland, Y.-S. Lee, N. Tjandra, R. A. Byrd, R. E. Swenson, O. Vasalatiy, *Dalton Trans.*, **2016**, *45*, 4673-4687.
174. D. Parker, R. S. Dickins, H. Puschmann, C. Crossland, J. A. K. Howard, *Chem. Rev.*, **2002**, *102*, 1977-2010.
175. E. A. Sutura, D. Häussinger, K. Zimmermann, L. Garbuio, M. Yulikov, G. Jeschke, I. Kuprov, *Chem. Sci.*, **2017**, *8*, 2751-2757.
176. B. Graham, C. T. Loh, J. D. Swarbrick, P. Ung, J. Shin, H. Yagi, X. Jia, S. Chhabra, N. Barlow, G. Pintacuda, T. Huber, G. Otting, *Bioconjugate Chem.*, **2011**, *22*, 2118-2125.
177. F. Yang, X. Wang, B.-B. Pan, X.-C. Su, *Chem. Commun.*, **2016**, *52*, 11535-11538.

178. W.-M. Liu, P. H. J. Keizers, M. A. S. Hass, A. Blok, M. Timmer, A. J. C. Sarris, M. Overhand, M. Ubbink, *J. Am. Chem. Soc.*, **2012**, *134*, 17306-17313.
179. W.-M. Liu, S. P. Skinner, M. Timmer, A. Blok, M. A. S. Hass, D. V. Filippov, M. Overhand, M. Ubbink, *Chem. Eur. J.*, **2014**, *20*, 6256-6258.
180. L. Pauling, *Nature*, **1948**, *161*, 707-709.
181. L. Pauling, *Chem. Eng. News*, **1946**, *24*, 1375-1377.
182. E. Z. Eisenmesser, D. A. Bosco, M. Akke, D. Kern, *Science*, **2002**, *295*, 1520-1523.
183. K. A. Henzler-Wildman, V. Thai, M. Lei, M. Ott, M. Wolf-Watz, T. Fenn, E. Pozharski, M. A. Wilson, G. A. Petsko, M. Karplus, C. G. Hübner, D. Kern, *Nature*, **2007**, *450*, 838-844.
184. L. Cao, O. Caldararu, A. C. Rosenzweig, U. Ryde, *Angew. Chem. Int. Ed.*, **2018**, *57*, 162-166.
185. K. E. H. Frandsen, T. J. Simmons, P. Dupree, J.-C. N. Poulsen, G. R. Hemsworth, L. Ciano, E. M. Johnston, M. Tovborg, K. S. Johansen, P. von Freiesleben, L. Marmuse, S. Fort, S. Cottaz, H. Driguez, B. Henrissat, N. Lenfant, F. Tuna, A. Baldansuren, G. J. Davies, L. Lo Leggio, P. H. Walton, *Nat. Chem. Biol.*, **2016**, *12*, 298-303.
186. S. T. Prigge, A. S. Kolhekar, B. A. Eipper, R. E. Mains, L. M. Amzel, *Nat. Struct. Biol.*, **1999**, *6*, 976-983.
187. T. V. Vendelboe, P. Harris, Y. Zhao, T. S. Walter, K. Harlos, K. El Omari, H. E. M. Christensen, *Sci. Adv.*, **2016**, *2*, e1500980.
188. C. M. Wilmot, J. Hajdu, M. J. McPherson, P. F. Knowles, S. E. V. Phillips, *Science*, **1999**, *286*, 1724-1728.
189. A. Changela, K. Chen, Y. Xue, J. Holschen, C. E. Outten, T. V. Halloran, A. Mondragón, *Science*, **2003**, *301*, 1383.
190. A. K. Wernimont, D. L. Huffman, A. L. Lamb, T. V. O'Halloran, A. C. Rosenzweig, *Nat. Struct. Biol.*, **2000**, *7*, 766-771.
191. M. Meury, M. Knop, F. P. Seebeck, *Angew. Chem. Int. Ed.*, **2017**, *56*, 8115-8119.
192. K. Watanabe, T. Masuda, H. Ohashi, H. Mihara, Y. Suzuki, *Eur. J. Biochem.*, **1994**, *226*, 277-283.
193. M. Knop, P. Engi, R. Lemnar, F. P. Seebeck, *ChemBioChem*, **2015**, *16*, 2147-2150.
194. M. Knop, T. Q. Dang, G. Jeschke, F. P. Seebeck, *ChemBioChem*, **2017**, *18*, 161-165.
195. B. L. Carlson, E. R. Ballister, E. Skordalakes, D. S. King, M. A. Breidenbach, S. A. Gilmore, J. M. Berger, C. R. Bertozzi, *J. Biol. Chem.*, **2008**, *283*, 20117-20125.
196. J. H. Austin, *JAMA Neurol.*, **1973**, *28*, 258-264.
197. G. Diez-Roux, A. Ballabio, *Annu. Rev. Genomics Hum. Genet.*, **2005**, *6*, 355-379.
198. M. J. Appel, C. R. Bertozzi, *ACS Chem. Biol.*, **2015**, *10*, 72-84.
199. P. Bojarová, S. J. Williams, *Curr. Opin. Chem. Biol.*, **2008**, *12*, 573-581.
200. M. J. Appel, K. K. Meier, J. Lafrance-Vanasse, H. Lim, C.-L. Tsai, B. Hedman, K. O. Hodgson, J. A. Tainer, E. I. Solomon, C. R. Bertozzi, *Proc. Natl. Acad. Sci.*, **2019**, *116*, 5370-5375.
201. A. D. Waldman, J. M. Fritz, M. J. Lenardo, *Nat. Rev. Immunol.*, **2020**, *20*, 651-668.
202. M. Lepore, L. Mori, G. De Libero, *Front. Immunol.*, **2018**, *9*, 1365.
203. J. E. Klaunig, L. M. Kamendulis, *Annu. Rev. Pharmacol. Toxicol.*, **2004**, *44*, 239-267.
204. L. J. Marnett, *Mutat. Res.*, **1999**, *424*, 83-95.
205. L. J. Marnett, J. N. Riggins, J. D. West, *J. Clin. Invest.*, **2003**, *111*, 583-593.
206. J. Szekely, H. Wang, K. M. Peplowski, C. G. Knutson, L. J. Marnett, C. J. Rizzo, *Nucleosides Nucleotides Nucleic Acids*, **2008**, *27*, 103-109.

3 Part I

3.1 Conformationally locked lanthanide chelating tags for convenient pseudocontact shift protein nuclear magnetic resonance spectroscopy



Conformationally locked lanthanide chelating tags for convenient pseudocontact shift protein nuclear magnetic resonance spectroscopy

Daniel Joss¹ · Roché M. Walliser¹ · Kaspar Zimmermann¹ · Daniel Häussinger¹

Received: 4 June 2018 / Accepted: 9 August 2018 / Published online: 16 August 2018
© Springer Nature B.V. 2018

Abstract

Pseudocontact shifts (PCS) generated by lanthanide chelating tags yield valuable restraints for investigating protein structures, dynamics and interactions in solution. In this work, dysprosium-, thulium- and terbium-complexes of eight-fold methylated 1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid tags [DOTA-M8-(4R4S)-SSPy] are presented that induce large pseudocontact shifts up to 5.5 ppm and adopt exclusively the square antiprismatic conformation. This is in contrast to our earlier findings on complexes of the stereoisomeric DOTA-M8-(8S)-SSPy, where significant amounts of the twisted square antiprismatic conformer for the Dy tag were observed. The Dy-, Tm-, Tb- and Lu-complexes of DOTA-M8-(4R4S)-SSPy were conjugated to ubiquitin S57C and selectively ¹⁵N leucine labeled human carbonic anhydrase II S50C, resulting in only one set of signals. Furthermore, we investigated the conformation of the thulium- and dysprosium-complexes in vacuo and with implicit water solvent using density functional theory calculations. The calculated energy differences between the two different conformations (7.0–50.5 kJ/mol) and experimental evidence from the corresponding ytterbium- and yttrium-complexes clearly suggest a SAP [$\Lambda(\delta\delta\delta\delta)$] geometry for the complexes presented in this study. The lanthanide chelating tag studied in this work offer insights into the solution structure of proteins by inducing strong pseudocontact shifts, show different tensor properties compared to its predecessor, enables a convenient assignment procedure, is accessed by a more economic synthesis than its predecessor and constitutes a highly promising starting point for further developments of lanthanide chelating tags.

Keywords Nuclear magnetic resonance · Pseudocontact shift · Lanthanide chelating tag · Paramagnetic · Density functional theory · Protein

Introduction

Pseudocontact shifts (PCS) generated by lanthanide chelating tags yield valuable restraints for investigating protein structures, dynamics and interactions in solution (Gaponenko et al. 2002; Wöhnert et al. 2003; Ikegami et al. 2004; Pintacuda et al. 2006; Keizers et al. 2008, Su et al. 2008, 2009; Häussinger et al. 2009; Otting 2010; Peters et al. 2011; Loh et al. 2013; Liu et al. 2014; Brewer et al. 2015; Hikone et al. 2016; Müntener et al. 2016; Pan et al. 2016; Nitsche and Otting 2017). In order to access structural restraints by PCS measurements, rigidified 1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid (DOTA)-based chelators offer a promising scaffold (Ranganathan et al. 2002a, b; Häussinger et al. 2009). The methyl group substituents on the nitrogen containing macrocycle adopt an equatorial-upper position (Parker et al. 2002; Opina et al. 2016) when the ligand is coordinated to a lanthanide metal ion and prevent motional averaging and

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10858-018-0203-4>) contains supplementary material, which is available to authorized users.

✉ Daniel Häussinger
daniel.haeussinger@unibas.ch

Daniel Joss
daniel.joss@unibas.ch

Roché M. Walliser
roche.walliser@gmail.com

Kaspar Zimmermann
zimkas00@gmail.com

¹ Department of Chemistry, University of Basel, St. Johanns-Ring 19, 4056 Basel, Switzerland

line broadening of signals by locking the 12-membered ring in a ($\delta\delta\delta\delta$) conformation (Häussinger et al. 2009).

However, Ln-DOTA-M8-(8S)-SSPy can show two signal sets, which results in a complicated analysis of spectra obtained by nuclear magnetic resonance spectroscopy containing the structural restraints for protein structure determination. As shown by Opina et al., the two signal sets have their origin in the presence of two different conformers of the lanthanide chelating tag, which can be interconverted to each other by rotation of the pendant arms of the chelator (Opina et al. 2016). In the course of their investigations, it was shown that for the Ln-DOTA-M8-(8S)-SSPy lanthanide chelating tag both arm rotation conformers are present in significant ratios for lanthanides heavier than europium and lighter than ytterbium and that the change in the ratio of the two possible conformers is dependent on the lanthanide ionic radii and therefore ultimately caused by the lanthanide contraction. Furthermore, it was demonstrated that the two conformers are in slow exchange on a timescale of approximately 4 h (Opina et al. 2016).

Especially for the strongly paramagnetic lanthanides dysprosium and thulium, it would be highly favorable to have only one conformational species present in rigidified DOTA-derived lanthanide chelating tags. We supposed that the pendant arms can be locked in one conformation by inverting the configuration of the stereocenter of the pendant arms [DOTA-M8-(4R4S)-SSPy], since the methyl substituents on the pendant arms would avoid steric clashes with the methyl substituents from the 12-membered ring.

Besides the locked conformation, we also expected different tensor properties compared to the DOTA-M8-(8S)-SSPy tag, because it was suggested by Mironov et al. that small changes in the coordination polyhedron can significantly affect the tensor properties (Mironov et al. 2001). In order to investigate this approach, we synthesized Dy-, Tm-, Tb- and Lu-DOTA-M8-(4R4S)-SSPy tags (Fig. 1) and conjugated each lanthanide complex to ubiquitin S57C and to selectively ^{15}N leucine labeled human carbonic anhydrase II S50C protein constructs. The obtained spectra as well as

tensor sizes and orientations relative to the protein were analyzed.

Density functional theory calculations were envisioned to assign the conformation of the DOTA-M8-(4R4S)-SSPy tags, since 2D NMR spectra as for example used by Ranganathan et al. for the assignment of Yb^{3+} M4-DOTMA chelates (Ranganathan et al. 2002a, b) are difficult or even impossible to measure for the two strongly paramagnetic lanthanide chelating tags due the extremely short T2 relaxation time caused by the close paramagnetic center (Helm 2006). Density functional theory calculations gained recently attraction to study lanthanide complexes and especially their conformational behavior, since they present a highly accurate but still fast and reliable quantum mechanical approach for structure calculation of lanthanide chelating tags (Cosentino et al. 2002; Natrajan et al. 2010; Regueiro-Figueroa et al. 2011; Blahut et al. 2017). Besides the fact that suitable effective core potentials are available for the lanthanide series, e.g. the ones reported by Dolg et al. (1989, 1993), the electrons can also be treated explicitly and including relativistic effects provided sufficiently powerful computational infrastructure is available. DFT calculations are an important tool for geometry optimization studies, calculation of relative energies and other molecular properties of both common organic molecules and metal complexes (Cramer and Truhlar 2009). We therefore compared the energies of SAP and TSAP geometry for the thulium and dysprosium complex and assigned the most stable conformation in aqueous solution.

Materials and methods

Synthesis

Lanthanide complexes were synthesized according to the procedures reported by Häussinger et al. for the Lu-DOTA-M8-(4R4S)-SSPy tag (Häussinger et al. 2009). In the amide coupling reaction, PyBOP (benzotriazol-1-yl-oxypyrrolidinophosphonium hexafluorophosphate) was replaced

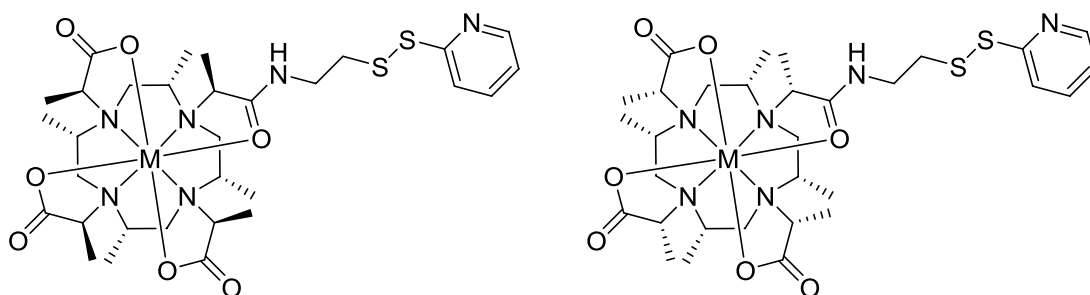


Fig. 1 DOTA-M8-(8S)-SSPy (left) and DOTA-M8-(4R4S)-SSPy (right)

by HATU (1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate) resulting in a higher yield of about 91% for this particular step in the synthesis. The side arms of DOTA-M8-(4*R*4*S*)-SSPy can conveniently be obtained from L-lactic acid, compared to the more expensive D-lactic acid needed for the synthesis of the DOTA-M8-(8*S*)-SSPy tag.

Expression of ubiquitin S57C and human carbonic anhydrase II S50C mutants and tagging reaction

Ubiquitin S57C was expressed as described previously by Sass et al. (1999) selectively ¹⁵N leucine labeled hCA II S50C as described by Varghese et al. (2016). The tagging reactions were carried out accordingly to the procedure reported by Häussinger et al. (2009), except that the pH during tagging was kept at pH 7.0 for ubiquitin S57C and pH 6.8 for hCA II S50C to ensure a faster tagging reaction.

NMR experiments and determination of the paramagnetic susceptibility tensor

¹H, ROESY, COSY, HMQC and HMBC spectra were measured in D₂O at a temperature of 298 K on 500 and 600 MHz Bruker Avance III NMR spectrometers. ¹H-¹⁵N HSQC spectra were measured in 10 mM phosphate buffer with pH 6.0 (ubiquitin S57C) and pH 6.8 (hCA II S50C) at a temperature of 298 K on a 600 MHz Bruker Avance III NMR spectrometer equipped with a cryogenic QCI probe. The obtained NMR spectra were assigned using CcpNmr Analysis (Vranken et al. 2005). The tensor properties were then obtained by fitting to the secondary structure elements of ubiquitin [PDB 1UBI (Ramage et al. 1994)] or the leucine residues of hCA II [PDB 3KS3 (Avvaru et al. 2010)] using Numbat (Schmitz et al. 2008). Q-factors were calculated using the following equation:

$$Q = \frac{\sqrt{\sum (\text{PCS}_{\text{exp}} - \text{PCS}_{\text{calc}})^2}}{\sqrt{\sum (\text{PCS}_{\text{exp}})^2}}$$

DFT calculations

DFT calculations were performed with the ORCA program package (Neese 2012) at the sciCORE facility of the University of Basel. For the calculations, BP86 was used as functional, SARC-TZVP as basis set for the ligands, while SARC2-QZVP was used as basis set for the lanthanide metal. The calculations were performed using the relativistic ZORA approximation, as well as the RI approximation to speed up the calculations. To model the water solvent, CPCM solvent model was implemented into the calculations.

Results and discussion

In order to access a lanthanide chelating tag with a completely rigidified and conformationally locked scaffold that yields only one set of signals when coupled to the protein of interest and exhibit novel tensor properties, we inverted the stereochemistry of the pendant arms of the DOTA-M8-(8*S*)-SSPy tag giving DOTA-M8-(4*R*4*S*)-SSPy. The previous findings by Opina et al. (2016), i.e. the formation of two conformers for the dysprosium and thulium tag resulting in two NMR signal sets for the measured protein, were not found in our case for the dysprosium-, thulium-, terbium- and lutetium metal-complexes when attached to a protein scaffold (Ln-DOTA-M8-(4*R*4*S*)-Ub^{S57C}). The four lanthanides were chosen because they displayed high (Dy, 27%; Tb, 20%), low (Tm, 9%) and negligible amounts (Lu, 2%) of the 2nd conformer in Ln-DOTA-M8-(8*S*)-SSPy and because of their use as diamagnetic reference (lutetium) and the valuable PCS that they generate (dysprosium, thulium and terbium).

Identification of one conformational species, attachment to ubiquitin S57C and tensor sizes and properties

The described tag shows only one species for dysprosium-, thulium- and terbium- and lutetium-complexes in the crude product, during HPLC purification and after storage for several weeks (HPLC-ESI-MS traces in supporting information, Figs. S1, S2, S3, S4) and only one set of signals in HSQC spectra when attached to ubiquitin S57C (Fig. 2) or human carbonic anhydrase II S50C (Fig. 3).

Compared to DOTA-M8-(8*S*)-SSPy, the ligand in the present study displays altered tensor properties (Table 1). The thulium complex offers properties well suited for the 3D structure refinement of proteins, since the large axial component results in valuable PCS (Fig. 4). The obtained pseudocontact shifts using the thulium complex are in the range of current high-performance lanthanide chelating tags (Keizers et al. 2008; Häussinger et al. 2009; Graham et al. 2011; Peters et al. 2011; Müntener et al. 2016; Chen et al. 2018).

The pseudocontact shifts obtained for the dysprosium complex are in a range that ensures valuable structural restraints. Furthermore, the assignment is facilitated by encountering only one set of signals. To our surprise, the dysprosium chelating tag yields smaller PCS values than the thulium tag, as a consequence of the smaller tensor parameters. This is in strong contrast to the results with Ln-DOTA-M8-(8*S*)-Ub^{S57C}, where the dysprosium tag showed a paramagnetic susceptibility tensor with

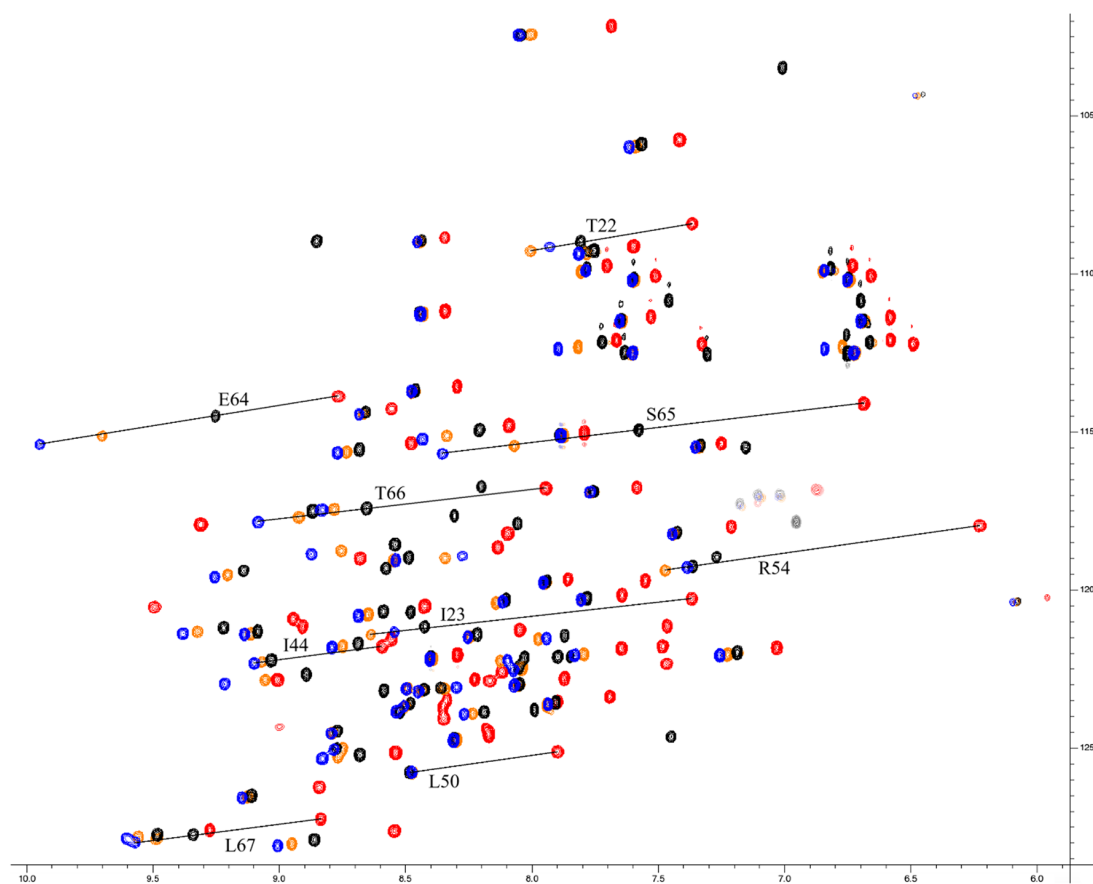


Fig. 2 Overlay of ^1H - ^{15}N HSQC spectra of Dy- (blue), Tm- (red), Tb- (orange) and Lu-DOTA-M8-(4R4S)-Ub $^{\text{S57C}}$ (black)

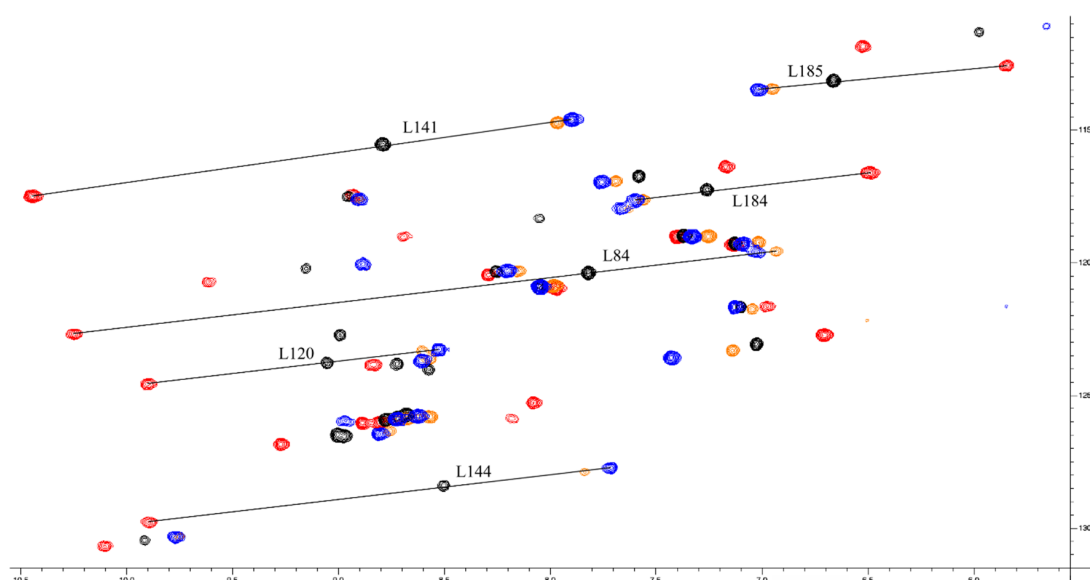
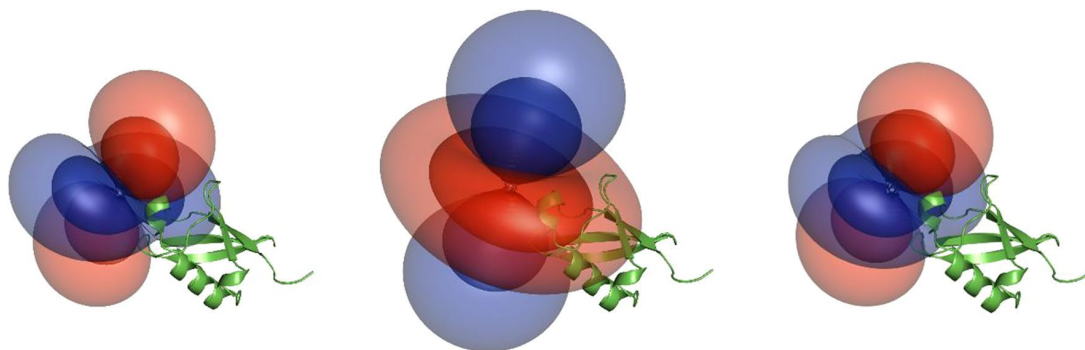


Fig. 3 Overlay of ^1H - ^{15}N HSQC spectra of Dy- (blue), Tm- (red), Tb- (orange) and Lu-DOTA-M8-(4R4S)-SSPy (black) attached to selectively ^{15}N leucine labeled hCA II S50C

Table 1 Properties of the induced paramagnetic susceptibility tensors measured on ubiquitin S57C at 298 K and pH 6.0

# PCS	Ln ³⁺	$\Delta\chi_{ax}$ (10 ⁻³² m ³)	$\Delta\chi_{rh}$ (10 ⁻³² m ³)	X _{metal}	Y _{metal}	Z _{metal}	α	β	γ	Q (%)
74	Dy	-8.7	-4.1	16.8	14.6	9.0	160.0	37.6	37.2	13.6
68	Tm	19.6	3.0	16.8	14.6	9.0	174.7	44.9	83.9	6.2
74	Tb	-8.2	-3.3	16.8	14.6	9.0	158.6	36.3	53.7	20.1

**Fig. 4** The tensors generated by the dysprosium (left), thulium (middle) and terbium complex (right) and their relative orientation to ubiquitin S57C [PCS isosurfaces set to 1 ppm (inner layer) and 0.25 ppm (outer layer)]

considerably higher values for $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$. When attached to ubiquitin S57C, the terbium complex yields tensor properties comparable to the dysprosium complex. The obtained results for Dy-DOTA-M8-(4R4S)-Ub^{S57C} could be explained by the different orientation of the sidearms of the lanthanide chelating tag [SAP or $\Lambda(\delta\delta\delta\delta)$] compared to Dy-DOTA-M8-(8S)-Ub^{S57C} [TSAP or $\Delta(\delta\delta\delta\delta)$]. Not only are the oxygens of the carboxy groups in the (4R4S)-tag placed in the plane above the basic ring between the nitrogen atoms of the macrocycle, furthermore the dipoles of the C–O bonds of the carboxy group point in a different direction. These electronic and positional changes create a strongly different environment for the lanthanide metal and thereby significantly affect the PCS values and tensor properties as predicted by Mironov et al. (2001). Discrepancies between the classic theory of Bleaney and experimental findings are subject of current research (Bleaney 1972; Funk et al. 2015; Sutturina et al. 2017).

Surprised by the smaller tensor generated by the dysprosium complex and driven by scientific curiosity to further benchmark our lanthanide chelating tag, we decided to further validate our results on hCA II S50C.

Benchmarking on selectively ¹⁵N leucine labeled human carbonic anhydrase II S50C

In order to verify and confirm the tensor properties obtained for ubiquitin S57C, we tagged selectively ¹⁵N leucine labeled

hCA II S50C, a 29 kDa protein construct, with the dysprosium-, thulium-, terbium- and lutetium-complexes.

The results show that the paramagnetic susceptibility tensor generated on hCA II S50C is bigger than the tensor for ubiquitin S57C (Fig. 5; Table 2). This arises most likely from a more restricted translational and/or rotational motion of the tag on the human carbonic anhydrase II. Whereas in our case the higher mobility of the spin label decreases the observed anisotropy tensor, also the opposite effect was described by Shishmarev et al. for very mobile spin labels that arises from application of the fitting procedure with an apparent positional deviation of the metal away from the proteins surface due to the motion of the tag thus enhancing the tensor magnitude (Shishmarev and Otting 2013). Both obtained tensor shapes for hCA II S50C and their relative magnitude when compared to each other remained the same as for ubiquitin S57C.

The orientation of all investigated complexes with regard to the protein could also render them ideally suitable to obtain structural restraints for studies of a protein dimer complex, in which the second protein would bind in the region where the paramagnetic susceptibility tensor of the lanthanide chelating tag has its maximum impact (Figs. 4, 5).

Assignment of the conformation of Dy- and Tm-DOTA-M8-(4R4S)-SSPy by DFT calculations and determination of structures and stabilization energies of the two conformational species

Ranganathan et al. determined the Yb³⁺ M4-DOTMA-(4R4S) chelates to adopt SAP geometry using NMR

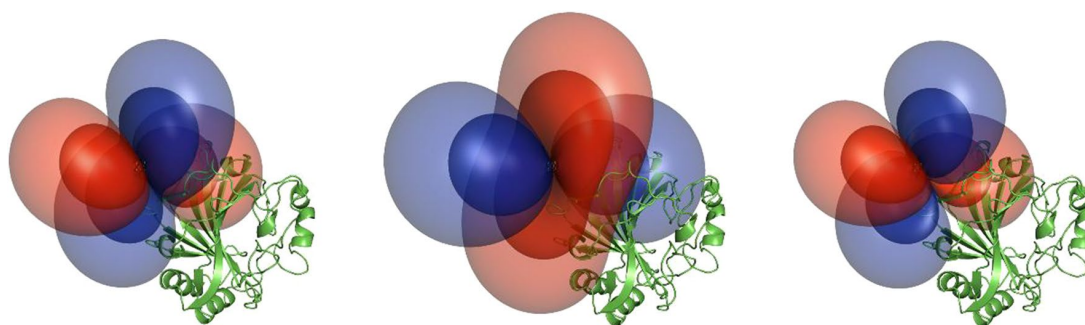


Fig. 5 The tensors generated by the dysprosium (left), thulium (middle) and the terbium complex (right) and their relative orientation to human carbonic anhydrase S50C [PCS isosurfaces set to 1.4 ppm (inner layer) and 0.35 ppm (outer layer)]

Table 2 Properties of the induced paramagnetic susceptibility tensors on hCA II S50C at 298 K and pH 6.8

# PCS	Ln ³⁺	$\Delta\chi_{ax}$ (10^{-32} m ³)	$\Delta\chi_{rh}$ (10^{-32} m ³)	X _{metal}	Y _{metal}	Z _{metal}	α	β	γ	Q (%)
46	Dy	-21.0	-11.1	-30.1	13.1	21.4	5.6	99.1	74.4	13.0
46	Tm	36.1	11.5	-30.1	13.1	21.4	179.0	85.1	117.8	7.0
42	Tb	-16.7	-11.8	-30.1	13.1	21.4	2.2	99.8	66.5	21.5

experiments and molecular mechanics (Ranganathan et al. 2002a, b). The experimental data and the structural assignments for the Yb³⁺ M4-DOTMA-(4R4S) chelate determined by Ranganathan et al. (2002a, b) and our assignment for Lu-DOTA-M8-(4R4S)-SSPy (Häussinger et al. 2009), provided a good starting point for our conformational analysis. We performed DFT calculations to estimate the energy difference between SAP and TSAP conformers and to assign the more stable conformation for the dysprosium and thulium complex, since measuring 2D spectra of these two complexes poses insurmountable difficulties due to the extremely short T2 relaxation time caused by the close paramagnetic center (Helm 2006). The DFT calculations were performed using the BP86 functional of Becke and Perdew (Perdew 1986; Becke 1988), which shows reliable and accurate results for the comparison of energies since the errors generated by this functional occur in a mainly systematic manner, in contrast to other functionals containing empirical parameters. For example B3LYP, which can be considered as one of the most popular functionals for small organic molecules with regard to obtained results and time and effort needed (Sousa et al. 2007), was shown to perform worse than GGA functionals for lanthanide complexes and introduces increasing errors for big molecules (Redfern et al. 2000; Grimme 2005; Grimme et al. 2016). The calculations were performed in vacuum as well as with implicit water using the CPCM solvent model, which is known to give reliable and accurate results for solvents of high polarity/dielectric constant with deviations less than 1 kcal/mol for the aqueous solvation energy of neutral molecules at B3LYP level (Takano and Houk 2005).

Interestingly, the calculations show a stabilization towards the SAP conformer for all three metal complexes, with and without inner-shell water molecule, in vacuo and in water modeled with an implicit solvent model, with an energetic difference between the two conformers in the range of 7.0–50.5 kJ/mol (Table 3). The biggest stabilization towards the SAP isomer for an eight-coordinated complex in solution is obtained for the lutetium complex (25.0 kJ/mol, K = ~24,000). The calculated stabilization energies for the thulium- and dysprosium-complexes are smaller (10.0 kJ/mol and 10.7 kJ/mol), but still significantly exceed the thermal energy at 298 K and constitute an equilibrium constant of K = 55 and K = 76 towards the SAP conformer. The DFT optimized structures of the SAP conformers of the lutetium-, thulium- and dysprosium-complexes in vacuo are depicted in Fig. 6.

Interestingly, we did not obtain converged and meaningful 9-coordinated structures for the SAP lutetium- and thulium-complexes with an inner-shell water molecule in using the CPCM solvent model. However, we observed that the inner-shell water molecule was lost to the outer-shell and aligned in a way that hydrogen bonds can be formed with the carboxylates (Fig. 7). These findings from our calculations can be explained by the fact that for the smaller lanthanides ions, as thulium and lutetium, the eight-coordinated species without apical water molecule is preferred due to sterics (Benetollo et al. 2003; Strickland et al. 2016). Without solvent model (in vacuo), the water molecule stays close to the metal ion, presumably due to the interaction of the oxygen atom with the positively charged metal ion. However, using a solvent model, the water molecule is more favored to align distant

Table 3 Stabilization energies obtained with or without coordinated water molecule in vacuo and water solvent [L=DOTA-M8-(4R4S)-SSPy]

Calculated energies of the different complexes M(L) (E_H /Hartree)		Stabilization (kJ/mol)
Lu(L) TSAP – 16713.85773	Lu(L) SAP – 16713.87043	– 33.3
Lu(L) TSAP solv. – 16713.95067	Lu(L) SAP solv. – 16713.96020	– 25.0
Lu(L)-H ₂ O TSAP – 16790.34318	Lu(L)-H ₂ O SAP – 16790.36243	– 50.5
Lu(L)-H ₂ O TSAP solv. – 16790.42518	Lu(L)-H ₂ O SAP solv. – ^a	–
Tm(L) TSAP – 15802.89887	Tm(L) SAP – 15802.90695	– 21.2
Tm(L) TSAP solv. – 15802.98688	Tm(L) SAP solv. – 15802.99067	– 10.0
Tm(L)-H ₂ O TSAP – 15879.38336	Tm(L)-H ₂ O SAP – 15879.39700	– 35.8
Tm(L)-H ₂ O TSAP solv. – ^a	Tm(L)-H ₂ O SAP solv. – ^a	–
Dy(L) TSAP – 14504.81886	Dy(L) SAP – 14504.82753	– 22.8
Dy(L) TSAP solv. – 14504.90612	Dy(L) SAP solv. – 14504.91021	– 10.7
Dy(L)-H ₂ O TSAP – 14581.30633	Dy(L)-H ₂ O SAP – 14581.31911	– 33.6
Dy(L)-H ₂ O TSAP solv. – 14581.38820	Dy(L)-H ₂ O SAP solv. – 14581.39086	– 7.0

^a–: no converged, meaningful, nine-coordinated complex obtained in the performed DFT calculations

from the metal ion and the interaction of the oxygen atom with the metal ion is no longer beneficial for stabilization.

The SAP geometry found by Ranganathan et al. for the Yb³⁺ M4-DOTMA-(4R4S) chelates and our assignment for Lu-DOTA-M8-(4R4S)-SSPy (Häussinger et al. 2009) combined with the results of our DFT calculations clearly suggest that the dysprosium and thulium complexes presented

in this work adopt a SAP [$\Lambda(\delta\delta\delta\delta)$] geometry. DFT calculations using other functionals and more expensive reference calculations are in progress and will be published in a separate computational study that benchmarks the accuracy and suitability of current popular and inexpensive DFT functionals used for structure and energy calculations of lanthanide complexes. To get further insights and validate our computational results, we decided to investigate the complexes by additional experimental studies.

Experimental validation of the conformational bias obtained in DFT calculations

In order to verify the conformational bias obtained in the performed DFT calculations, we recorded ¹H spectra of the dysprosium-, thulium-, terbium-, ytterbium- and yttrium-complexes (Figs. S8, S9, S10, S11, S12), fitted the obtained pseudocontact shifts for the ytterbium complex to the calculated structures of the corresponding SAP and TSAP conformers and analyzed the diamagnetic yttrium complex that shows an ionic radius for 3+ charged, eight-coordinated species that lies between the radius obtained for the corresponding dysprosium and thulium ions using 2D NMR experiments (Shannon 1976).

The synthesized ytterbium complex was fitted according to the assignments by Ranganathan et al. (2002a, b) and the procedure applied by Gempf et al. (2013) to the calculated structures of the SAP and TSAP conformers of the ytterbium complex. The average shift of the axial protons closest to the metal center ($\delta_{PCS} = 216.7$ ppm), the ring methyl group protons ($\delta_{PCS} = 12.0$ ppm) and the methyl group protons of the acetate arms ($\delta_{PCS} = -33.9$ ppm) suggests clearly a SAP conformation (SAP: Q-factor 10.3% and $R^2 = 0.99$, TSAP: Q-factor 22.0% and $R^2 = 0.96$).

Based on the full assignment and the measured ROESY spectrum of the Y-DOTA-M8-(4R4S)-SSPy complex (Figs. S12, S13, S14, S15, S16), we assigned the adopted geometry to SAP [$\Lambda(\delta\delta\delta\delta)$]. The through space correlations detected from the α -protons of the carboxylate sidearms to the axial lower ring protons in the absence of through space correlations of the α -protons of the carboxylate sidearms to the

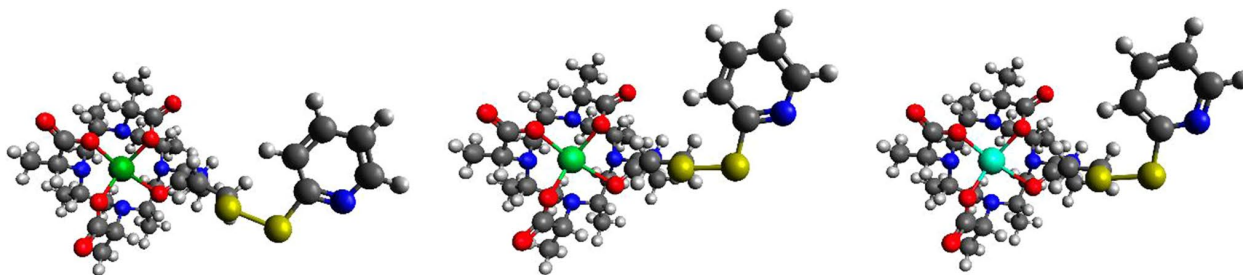
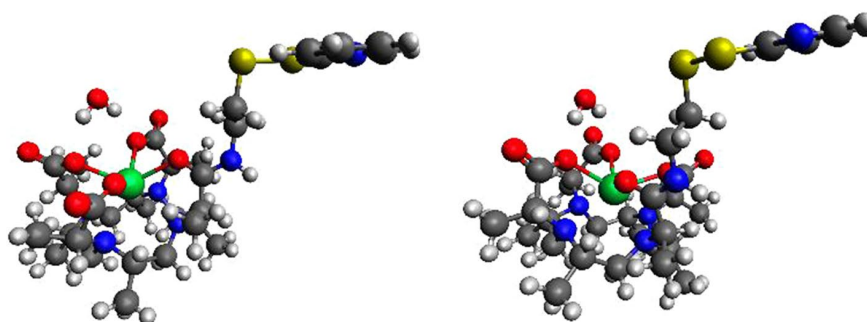
**Fig. 6** Structures of the SAP conformers of the lutetium-, thulium- and dysprosium-complexes in vacuo

Fig. 7 Eight-coordinated TSAP (left) and SAP (right) conformer of the Tm-DOTA-M8-(4*R*4*S*)-SSPy



equatorial lower ring protons clearly indicate a SAP geometry for the yttrium complex.

Based on the performed NMR experiments that suggest clearly a SAP geometry for the ytterbium- (ionic radius close to thulium) and yttrium-complexes (ionic radius between thulium and dysprosium) and the similar HPLC retention time of the analyzed complexes, we assigned the dysprosium- and thulium-complex to adopt a SAP [$\Lambda(\delta\delta\delta\delta)$] geometry.

Less expensive and higher yielding synthesis

Besides the different tensor properties and suitability for PCS measurements of the thulium and dysprosium ligands presented in this study compared to their predecessor, DOTA-M8-(8*S*)-SSPy, the less expensive and higher yielding synthesis presents an important step towards the more general and routine use of the DOTA-M8-SSPy lanthanide chelating tags for the structure determination of proteins.

Conclusion

To conclude, a strongly paramagnetic lanthanide chelating tag is reported that shows only one conformer and thus provide only one set of signals in HSQC experiments when attached to a protein. Furthermore, the investigated thulium- and dysprosium-complexes show different tensor properties when attached to ubiquitin S57C compared to their DOTA-M8-(8*S*)-SSPy analog. The geometry of the lanthanide chelating tag with thulium, dysprosium and lutetium in TSAP or SAP conformation, with and without inner-shell water molecule, in vacuo and in water modeled with a solvent model was calculated using DFT methods and the energetically most favorable conformation in all cases was assigned to be SAP. While the thulium complex is well suited for studies of the directly attached protein due to its tensor size and properties, both the dysprosium and thulium complex could be applied for study of a protein dimer, in which the second protein is oriented in a way that the yet unexploited tensor lobes maximally interact. Furthermore,

the comparison of Dy-DOTA-M8-(8*S*)-SSPy and Dy-DOTA-M8-(4*R*4*S*)-SSPy tag provides a starting point for further, more detailed studies on the interaction between geometry and tensor properties for lanthanide chelating tags. In order to enable studies on large proteins, protein complexes and other biopolymers, the development of high-performance lanthanide chelating tags has to be continued.

Acknowledgements The Chemistry Department of the University of Basel and the Swiss National Science Foundation Grant 200021_130263 are acknowledged for financial support. Calculations were performed at sciCORE (<http://scicore.unibas.ch/>) scientific computing core facility at University of Basel. Biological structures were generated using the open source software PyMOL (<http://www.pymol.org/>). C.E. Housecroft, E.C. Constable, T. Müntener and R. Vogel are acknowledged for helpful discussions. R.A. Byrd is gratefully acknowledged for a gift of M4-cyclen.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Avvaru BS, Kim CU, Sippel KH, Gruner SM, Agbandje-McKenna M, Silverman DN, McKenna R (2010) A short, strong hydrogen bond in the active site of human carbonic anhydrase II. *Biochemistry* 49:249–251
- Becke AD (1988) Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys Rev A* 38:3098–3100
- Benetollo F, Bombieri G, Calabi L, Aime S, Botta M (2003) Structural variations across the lanthanide series of macrocyclic DOTA complexes: insights into the design of contrast agents for magnetic resonance imaging. *Inorg Chem* 42:148–157
- Blahut J, Hermann P, Tosner Z, Platas-Iglesias C (2017) A combined NMR and DFT study of conformational dynamics in lanthanide complexes of macrocyclic DOTA-like ligands. *PCCP* 19:26662–26671
- Bleaney B (1972) Nuclear magnetic resonance shifts in solution due to lanthanide ions. *J Magn Reson* 8:91–100
- Brewer KD, Bacaj T, Cavalli A, Camilloni C, Swarbrick JD, Liu J, Zhou A, Zhou P, Barlow N, Xu J, Seven AB, Prinslow EA, Voleti R, Häussinger D, Bonvin AM, Tomchick DR, Vendruscolo M, Graham B, Südhof TC, Rizo J (2015) Dynamic binding

- mode of a Synaptotagmin-1-SNARE complex in solution. *Nat Struct Mol Biol* 22:555–564
- Chen JL, Zhao Y, Gong YJ, Pan B-B, Wang X, Su XC (2018) Stable and rigid DTPA-like paramagnetic tags suitable for in vitro and in situ protein NMR analysis. *J Biomol NMR* 70:77–92
- Cosentino U, Villa A, Pitea D, Moro G, Barone V, Maiocchi A (2002) Conformational characterization of lanthanide(III)-DOTA complexes by ab initio investigation in vacuo and in aqueous solution. *J Am Chem Soc* 124:4901–4909
- Cramer CJ, Truhlar DG (2009) Density functional theory for transition metals and transition metal chemistry. *PCCP* 11:10757–10816
- Dolg M, Stoll H, Savin A, Preuss H (1989) Energy-adjusted pseudopotentials for the rare earth elements. *Theor Chim Acta* 75:173–194
- Dolg M, Stoll H, Preuss H (1993) A combination of quasirelativistic pseudopotential and ligand field calculations for lanthanoid compounds. *Theor Chim Acta* 85:441–450
- Funk AM, Finney K, Harvey P, Kenwright AM, Neil ER, Rogers NJ, Kanthi Senanayake P, Parker D (2015) Critical analysis of the limitations of Bleaney's theory of magnetic anisotropy in paramagnetic lanthanide coordination complexes. *Chem Sci* 6:1655–1662
- Gaponenko V, Altieri AS, Li J, Byrd RA (2002) Breaking symmetry in the structure determination of (large) symmetric protein dimers. *J Biomol NMR* 24:143–148
- Gempfl KL, Butler SJ, Funk AM, Parker D (2013) Direct and selective tagging of cysteine residues in peptides and proteins with 4-nitropyridyl lanthanide complexes. *Chem Commun* 49:9104–9106
- Graham B, Loh CT, Swarbrick JD, Ung P, Shin J, Yagi H, Jia X, Chhabra S, Barlow N, Pintacuda G, Huber T, Otting G (2011) DOTA-amide lanthanide tag for reliable generation of pseudocontact shifts in protein NMR spectra. *Bioconj Chem* 22:2118–2125
- Grimme S (2005) Accurate calculation of the heats of formation for large main group compounds with spin-component scaled MP2 methods. *J Phys Chem A* 109:3067–3077
- Grimmel S, Schoendorff G, Wilson AK (2016) Gauging the performance of density functionals for lanthanide-containing molecules. *J Chem Theory Comput* 12:1259–1266
- Häussinger D, Huang JR, Grzesiek S (2009) DOTA-M8: an extremely rigid, high-affinity lanthanide chelating tag for PCS NMR spectroscopy. *J Am Chem Soc* 131:14761–14767
- Helm L (2006) Relaxivity in paramagnetic systems: theory and mechanisms. *Prog Nucl Magn Reson Spectrosc* 49:45–64
- Hikone Y, Hirai G, Mishima M, Inomata K, Ikeya T, Arai S, Shirakawa M, Sodeoka M, Ito Y (2016) A new carbamidomethyl-linked lanthanoid chelating tag for PCS NMR spectroscopy of proteins in living HeLa cells. *J Biomol NMR* 66:99–110
- Ikegami T, Verdier L, Sakhaei P, Grimme S, Pescatore B, Saxena K, Fiebig KM, Griesinger C (2004) Novel techniques for weak alignment of proteins in solution using chemical tags coordinating lanthanide ions. *J Biomol NMR* 29:339–349
- Keizers PHJ, Saragliadis A, Hiruma Y, Overhand M, Ubbink M (2008) Design, synthesis, and evaluation of a lanthanide chelating protein probe: CLaNP-5 yields predictable paramagnetic effects independent of environment. *J Am Chem Soc* 130:14802–14812
- Liu WM, Overhand M, Ubbink M (2014) The application of paramagnetic lanthanoid ions in NMR spectroscopy on proteins. *Coord Chem Rev* 273–274:2–12
- Loh CT, Ozawa K, Tuck KL, Barlow N, Huber T, Otting G, Graham B (2013) Lanthanide tags for site-specific ligation to an unnatural amino acid and generation of pseudocontact shifts in proteins. *Bioconj Chem* 24:260–268
- Mironov VS, Galyametdinov YG, Ceulemans A, Görller-Walrand C, Binnemans K (2001) Influence of crystal-field perturbations on the room-temperature magnetic anisotropy of lanthanide complexes. *Chem Phys Lett* 345:132–140
- Müntener T, Häussinger D, Selenko P, Theillet FX (2016) In-cell protein structures from 2D NMR experiments. *J Phys Chem Lett* 7:2821–2825
- Natrajan LS, Khoabane NM, Dadds BL, Muryn CA, Pritchard RG, Heath SL, Kenwright AM, Kuprov I, Faulkner S (2010) Probing the structure, conformation, and stereochemical exchange in a family of lanthanide complexes derived from tetrapyrrolyl-appended cyclen. *Inorg Chem* 49:7700–7709
- Neese F (2012) The ORCA program system. *Wiley Interdiscip Rev Comput Mol Sci* 2:73–78
- Nitsche C, Otting G (2017) Pseudocontact shifts in biomolecular NMR using paramagnetic metal tags. *Prog Nucl Magn Reson Spectrosc* 98–99:20–49
- Opina ACL, Strickland M, Lee YS, Tjandra N, Byrd RA, Swenson RE, Vasalatiy O (2016) Analysis of the isomer ratios of polymethylated-DOTA complexes and the implications on protein structural studies. *Dalton Trans* 45:4673–4687
- Otting G (2010) Protein NMR using paramagnetic ions. *Annu Rev Biophys* 39:387–405
- Pan BB, Yang F, Ye Y, Wu Q, Li C, Huber T, Su XC (2016) 3D structure determination of a protein in living cells using paramagnetic NMR spectroscopy. *Chem Commun* 52:10237–10240
- Parker D, Dickins RS, Puschmann H, Crossland C, Howard JAK (2002) Being excited by lanthanide coordination complexes: aqua species, chirality, excited-state chemistry, and exchange dynamics. *Chem Rev* 102:1977–2010
- Perdew JP (1986) Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys Rev B* 33:8822–8824
- Peters F, Maestre-Martinez M, Leonov A, Kovacic L, Becker S, Boelens R, Griesinger C (2011) Cys-Ph-TAHA: a lanthanide binding tag for RDC and PCS enhanced protein NMR. *J Biomol NMR* 51:329–337
- Pintacuda G, Park AY, Keniry MA, Dixon NE, Otting G (2006) Lanthanide labeling offers fast NMR approach to 3D structure determinations of protein-protein complexes. *J Am Chem Soc* 128:3696–3702
- Ramage R, Green J, Muir TW, Ogunjobi OM, Love S, Shaw K (1994) Synthetic, structural and biological studies of the ubiquitin system: the total chemical synthesis of ubiquitin. *Biochem J* 299:151–158
- Ranganathan RS, Pillai RK, Raju N, Fan H, Nguyen H, Tweedle MF, Desreux JF, Jacques V (2002a) Polymethylated DOTA ligands. 1. Synthesis of rigidified ligands and studies on the effects of alkyl substitution on acid-base properties and conformational mobility. *Inorg Chem* 41:6846–6855
- Ranganathan RS, Raju N, Fan H, Zhang X, Tweedle MF, Desreux JF, Jacques V (2002b) Polymethylated DOTA ligands. 2. Synthesis of rigidified lanthanide chelates and studies on the effect of alkyl substitution on conformational mobility and relaxivity. *Inorg Chem* 41:6856–6866
- Redfern PC, Zapol P, Curtiss LA, Raghavachari K (2000) Assessment of Gaussian-3 and density functional theories for enthalpies of formation of C1–C16 alkanes. *J Phys Chem A* 104:5850–5854
- Regueiro-Figueroa M, Bensenane B, Ruscsák E, Esteban-Gómez D, Charbonnière LJ, Tircsó G, Tóth I, Blas AD, Rodríguez-Blas T, Platas-Iglesias C (2011) Lanthanide DOTA-like complexes containing a picolinate pendant: structural entry for the design of Ln^{III}-based luminescent probes. *Inorg Chem* 50:4125–4141
- Sass J, Cordier F, Hoffmann A, Rogowski M, Cousin A, Omichinski JG, Löwen H, Grzesiek S (1999) Purple membrane induced alignment of biological macromolecules in the magnetic field. *J Am Chem Soc* 121:2047–2055
- Schmitz C, Stanton-Cook MJ, Su X-C, Otting G, Huber T (2008) Numbat: an interactive software tool for fitting $\Delta\chi$ -tensors to

- molecular coordinates using pseudocontact shifts. *J Biomol NMR* 41:179–189
- Shannon RD (1976) Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides. *Acta Cryst A* 32:751–767
- Shishmarev D, Otting G (2013) How reliable are pseudocontact shifts induced in proteins and ligands by mobile paramagnetic metal tags? A modelling study. *J Biomol NMR* 56:203–216
- Sousa SF, Fernandes PA, Ramos MJ (2007) General performance of density functionals. *J Phys Chem A* 111:10439–10452
- Strickland M, Schwieters CD, Göbl C, Opina ACL, Strub MP, Swenson RE, Vasalatiy O, Tjandra N (2016) Characterizing the magnetic susceptibility tensor of lanthanide-containing polymethylated-DOTA complexes. *J Biomol NMR* 66:125–139
- Su XC, Man B, Beeren S, Liang H, Simonsen S, Schmitz C, Huber T, Messerle BA, Otting G (2008) A dipicolinic acid tag for rigid lanthanide tagging of proteins and paramagnetic NMR spectroscopy. *J Am Chem Soc* 130:10486–10487
- Su XC, Liang H, Loscha KV, Otting G (2009) $[\text{Ln}(\text{DPA})_3]^{3-}$ is a convenient paramagnetic shift reagent for protein NMR studies. *J Am Chem Soc* 131:10352–10353
- Suturina EA, Mason K, Geraldes C, Kuprov I, Parker D (2017) Beyond Bleaney's theory: experimental and theoretical analysis of periodic trends in lanthanide-induced chemical shift. *Angew Chem Int Ed* 56:12215–12218
- Takano Y, Houk KN (2005) Benchmarking the conductor-like polarizable continuum model (CPCM) for aqueous solvation free energies of neutral and ionic organic molecules. *J Chem Theory Comput* 1:70–77
- Varghese S, Halling PJ, Häußinger D, Wimperis S (2016) High-resolution structural characterization of a heterogeneous biocatalyst using solid-state NMR. *J Phys Chem C* 120:28717–28726
- Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley JL, Ionides J, Laue ED (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins Struct, Funct, Bioinform* 59:687–696
- Wöhnert J, Franz KJ, Nitz M, Imperiali B, Schwalbe H (2003) Protein alignment by a coexpressed lanthanide-binding tag for the measurement of residual dipolar couplings. *J Am Chem Soc* 125:13338–13339

Supporting information

Conformationally locked lanthanide chelating tags for convenient pseudocontact shift protein nuclear magnetic resonance spectroscopy

Daniel Joss, MSc, Department of Chemistry, University of Basel, St. Johannis-Ring 19, CH-4056 Basel,
daniel.joss@unibas.ch, ORCID 0000-0002-9101-4525

Dr. Roché M. Walliser, Department of Chemistry, University of Basel, St. Johannis-Ring 19, CH-4056 Basel,
roche.walliser@gmail.com

Dr. Kaspar Zimmermann, Department of Chemistry, University of Basel, St. Johannis-Ring 19, CH-4056 Basel,
zimkas00@gmail.com

PD Dr. Daniel Häussinger, Head of NMR Laboratory, Department of Chemistry, University of Basel, St. Johannis-Ring 19, CH-4056 Basel, daniel.haeussinger@unibas.ch, ORCID 0000-0002-4798-0072

Analytical HPLC-ESI-MS measurement after several weeks of storage of dysprosium-, thulium-, lutetium-, terbium- and yttrium complexes

$\Lambda(\delta\delta\delta\delta)$ -Dysprosium-(2*R*,2'*R*,2''*R*)-2,2',2''-((2*S*,5*S*,8*S*,11*S*)-2,5,8,11-tetramethyl-10-((*R*)-1-oxo-1-(2-(pyridin-2-yl)disulfanyl)ethylamino)propan-2-yl)-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)tripropionate, $\Lambda(\delta\delta\delta\delta)$ -[Dy-DOTA-M8-(4*R*4*S*)-SSPy]

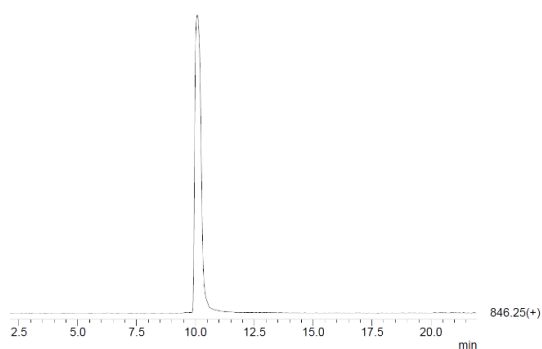


Figure S1: HPLC-ESI-MS trace of Dy-DOTA-M8-(4*R*4*S*)-SSPy.

$C_{31}H_{49}N_6DyO_7S_2$, M_{calc} ($[M+H]^+$, ^{162}Dy): 844.3, **ESI-MS**: m/z 843.3 (45.6%), 844.3 (78.6%), 845.3 (97.9%), 846.3 (100.0%), 847.3 (37.7%), 848.3 (14.2%).

$\Lambda(\delta\delta\delta\delta)$ -Thulium-(2*R*,2'*R*,2''*R*)-2,2',2''-((2*S*,5*S*,8*S*,11*S*)-2,5,8,11-tetramethyl-10-((*R*)-1-oxo-1-(2-(pyridin-2-yl)disulfanyl)ethylamino)propan-2-yl)-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)tripropionate, $\Lambda(\delta\delta\delta\delta)$ -[Tm-DOTA-M8-(4*R*4*S*)-SSPy]

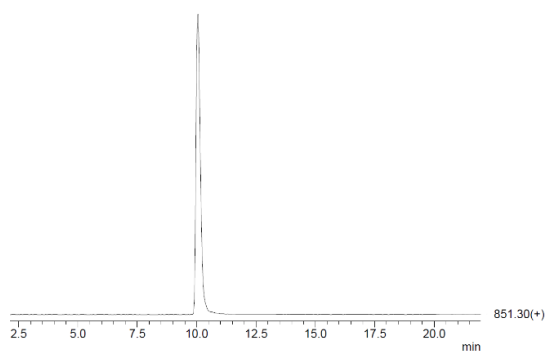


Figure S2: HPLC-ESI-MS trace of Tm-DOTA-M8-(4*R*4*S*)-SSPy.

$C_{31}H_{49}N_6TmO_7S_2$, M_{calc} ($[M+H]^+$): 851.3, **ESI-MS**: m/z 851.3 (100.0%), 852.3 (43.4%), 853.3 (18.3%).

$\Lambda(\delta\delta\delta\delta)$ -Lutetium-(2*R*,2'*R*,2''*R*)-2,2',2''-((2*S*,5*S*,8*S*,11*S*)-2,5,8,11-tetramethyl-10-((*R*)-1-oxo-1-(2-(pyridin-2-yl)disulfanyl)ethylamino)propan-2-yl)-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)tripropionate, $\Lambda(\delta\delta\delta\delta)$ -[Lu-DOTA-M8-(4*R*4*S*)-SSPy]

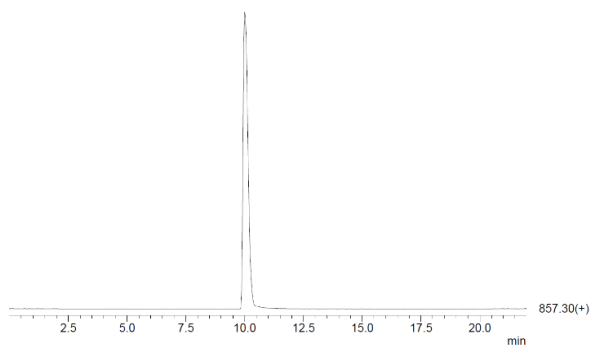


Figure S3: HPLC-ESI-MS trace of Lu-DOTA-M8-(4*R*4*S*)-SSPy.

$C_{31}H_{49}N_6LuO_7S_2$, $M_{calc}([M+H]^+)$: 857.3, **ESI-MS**: m/z 857.3 (100.0%), 858.3 (49.0%), 859.3 (20.1%).

$\Lambda(\delta\delta\delta\delta)$ -Terbium-(2*R*,2'*R*,2''*R*)-2,2',2''-((2*S*,5*S*,8*S*,11*S*)-2,5,8,11-tetramethyl-10-((*R*)-1-oxo-1-(2-(pyridin-2-yl)disulfanyl)ethylamino)propan-2-yl)-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)tripropionate, $\Lambda(\delta\delta\delta\delta)$ -[Tb-DOTA-M8-(4*R*4*S*)-SSPy]

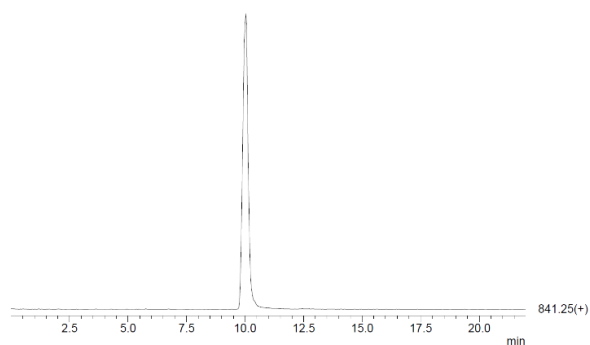


Figure S4: HPLC-ESI-MS trace of Tb-DOTA-M8-(4*R*4*S*)-SSPy.

$C_{31}H_{49}N_6TbO_7S_2$, $M_{calc}([M+H]^+)$: 841.3, **ESI-MS**: m/z 841.3 (100.0%), 842.3 (42.1%), 843.3 (19.0%).

$\Lambda(\delta\delta\delta\delta)$ -Yttrium-(2*R*,2'*R*,2''*R*)-2,2',2''-((2*S*,5*S*,8*S*,11*S*)-2,5,8,11-tetramethyl-10-((*R*)-1-oxo-1-(2-(pyridin-2-yl)disulfanyl)ethylamino)propan-2-yl)-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)tripropionate, $\Lambda(\delta\delta\delta\delta)$ -[Y-DOTA-M8-(4*R*4*S*)-SSPy]

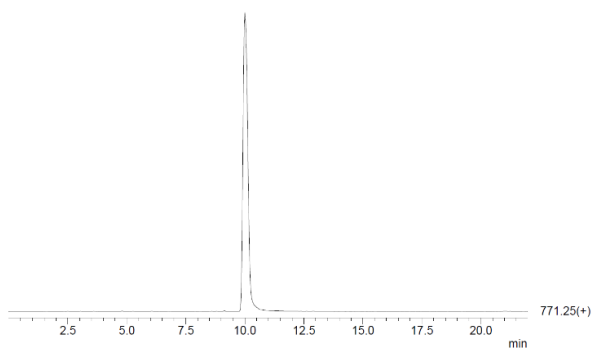


Figure S5: HPLC-ESI-MS trace of Y-DOTA-M8-(4*R*4*S*)-SSPy.

$C_{31}H_{49}N_6YO_7S_2$, $M_{\text{calc}}([M+H]^+)$: 771.2, **ESI-MS**: m/z 771.3 (100.0%), 772.2 (37.2%), 773.2 (17.1%).

Overlay of ^1H - ^{15}N HSQC spectra of Dy- (blue), Tm- (red), Tb- (orange) and Lu-DOTA-M8-(4R4S)-Ub S57C (black)

Measured in 10 mM phosphate buffer with pH 6.0 at a temperature of 298 K on a 600 MHz Bruker Avance III NMR spectrometer equipped with a cryogenic QCI probe.

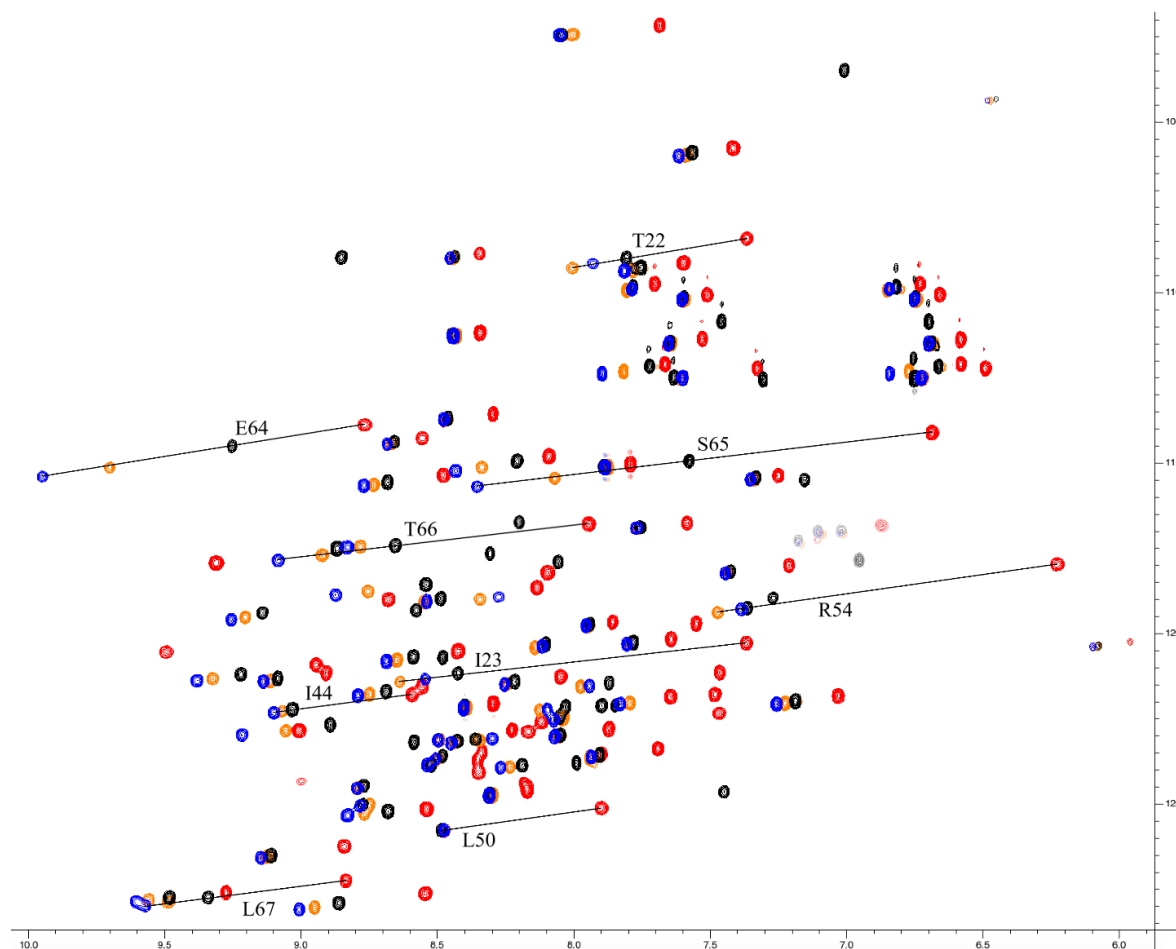


Figure S6: Overlay of ^1H - ^{15}N HSQC spectra of Dy- (blue), Tm- (red), Tb- (orange) and Lu-DOTA-M8-(4R4S)-Ub S57C (black).

Overlay of ^1H - ^{15}N HSQC spectra of Dy- (blue), Tm- (red), Tb- (orange) and Lu-DOTA-M8-(4R4S)-SSPy (black) attached to selectively ^{15}N leucine labeled human carbonic anhydrase II S50C

Measured in 10 mM phosphate buffer with pH 6.8 at a temperature of 298 K on a 600 MHz Bruker Avance III NMR spectrometer equipped with a cryogenic QCI probe.

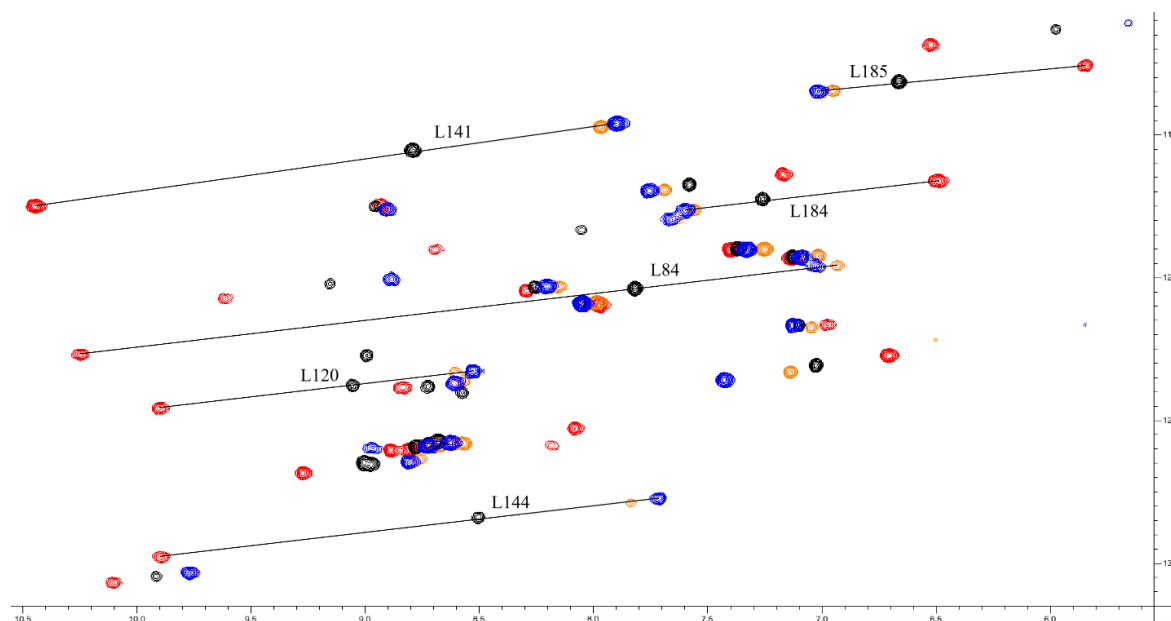


Figure S7: Overlay of ^1H - ^{15}N HSQC spectra of Dy- (blue), Tm- (red), Tb- (orange) and Lu-DOTA-M8-(4R4S)-SSPy (black) attached to selectively ^{15}N leucine labeled hCA II S50C.

Shift list comparison of ^1H - ^{15}N HSQC spectra of Dy- and Lu-DOTA-M8-(4R4S)-Ub^{S57C}

Table S1: Shift list comparison of ^1H - ^{15}N HSQC spectra of Dy- and Lu-DOTA-M8-(4R4S)-Ub^{S57C}.

Residue	Reson. 1	Shift Dy	Shift Lu	Reson. 2	Shift Dy	Shift Lu
2Gln	H	9.216	8.893	N	122.979	122.677
3Ile	H	8.432	8.207	N	115.233	114.942
4Phe	H	8.873	8.542	N	118.874	118.565
5Val	H	9.381	9.220	N	121.380	121.196
6Lys	H	9.007	8.861	N	128.094	127.906
7Thr	H	8.769	8.683	N	115.665	115.563
12Thr	H	8.687	8.588	N	120.819	120.685
13Ile	H	9.598	9.483	N	127.876	127.745
14Thr	H	8.792	8.688	N	121.822	121.702
15Leu	H	8.829	8.680	N	125.338	125.215
16Glu	H	8.074	8.051	N	122.529	122.456
23Ile	H	8.544	8.424	N	121.338	121.172
25Asn	H	7.942	7.871	N	121.537	121.453
26Val	H	8.096	8.030	N	122.239	122.156
27Lys	H	8.540	8.487	N	119.062	118.974
28Ala	H	7.937	7.907	N	123.616	123.569
29Lys	H	7.806	7.783	N	120.312	120.267
30Ile	H	8.255	8.217	N	121.499	121.413
31Gln	H	8.507	8.481	N	123.676	123.595
32Asp	H	7.956	7.942	N	119.775	119.729
33Lys	H	7.352	7.333	N	115.482	115.426
34Glu	H	8.684	8.659	N	114.451	114.377
39Asp	H	8.478	8.461	N	113.720	113.69
40Gln	H	7.771	7.757	N	116.905	116.882
41Gln	H	7.445	7.425	N	118.235	118.183
42Arg	H	8.450	8.425	N	123.219	123.153
43Leu	H	8.794	8.770	N	124.539	124.47
44Ile	H	9.099	9.031	N	122.316	122.229
45Phe	H	8.784	8.772	N	125.053	125.013
48Lys	H	7.831	7.897	N	122.068	122.12
49Gln	H	8.498	8.585	N	123.131	123.187
66Thr	H	9.084	8.654	N	117.852	117.424
67Leu	H	9.574	9.341	N	127.989	127.741
68His	H	9.256	9.141	N	119.595	119.394
69Leu	H	8.268	8.189	N	123.934	123.86
70Val	H	9.148	9.110	N	126.578	126.511
71Leu	H	8.071	8.048	N	123.033	122.985

Shift list comparison of ^1H - ^{15}N HSQC spectra of Tm- and Lu-DOTA-M8-(4R4S)-Ub^{S57C}

Table S2: Shift list comparison of ^1H - ^{15}N HSQC spectra of Tm- and Lu-DOTA-M8-(4R4S)-Ub^{S57C}.

Residue	Reson. 1	Shift Tm	Shift Lu	Reson. 2	Shift Tm	Shift Lu
2Gln	H	9.008	8.893	N	122.845	122.677
3Ile	H	8.092	8.207	N	114.806	114.942
4Phe	H	8.097	8.542	N	118.210	118.565
5Val	H	8.945	9.220	N	120.919	121.196
6Lys	H	8.545	8.861	N	127.631	127.906
7Thr	H	8.479	8.683	N	115.363	115.563
12Thr	H	8.424	8.588	N	120.516	120.685
13Ile	H	9.275	9.483	N	127.598	127.745
14Thr	H	8.556	8.688	N	121.550	121.702
15Leu	H	8.541	8.680	N	125.151	125.215
16Glu	H	8.119	8.051	N	122.578	122.456
23Ile	H	7.369	8.424	N	120.281	121.172
25Asn	H	7.466	7.871	N	121.149	121.453
26Val	H	7.646	8.030	N	121.855	122.156
27Lys	H	8.136	8.487	N	118.656	118.974
28Ala	H	7.693	7.907	N	123.380	123.569
29Lys	H	7.644	7.783	N	120.168	120.267
30Ile	H	8.049	8.217	N	121.269	121.413
32Asp	H	7.858	7.942	N	119.664	119.729
33Lys	H	7.251	7.333	N	115.374	115.426
34Glu	H	8.555	8.659	N	114.273	114.377
39Asp	H	8.296	8.461	N	113.567	113.69
40Gln	H	7.585	7.757	N	116.758	116.882
41Gln	H	7.212	7.45	N	117.998	118.183
42Arg	H	8.167	8.425	N	122.876	123.153
44Ile	H	8.592	9.031	N	121.790	122.229
48Lys	H	7.482	7.897	N	121.784	122.12
49Gln	H	8.227	8.585	N	122.830	123.187
66Thr	H	7.945	8.654	N	116.791	117.424
67Leu	H	8.836	9.341	N	127.246	127.741
68His	H	8.680	9.141	N	119.006	119.394
69Leu	H	7.901	8.189	N	123.533	123.86
70Val	H	8.841	9.110	N	126.244	126.511
71Leu	H	7.872	8.049	N	122.806	122.985

Shift list comparison of ^1H - ^{15}N HSQC spectra of Tb- and Lu-DOTA-M8-(4R4S)-Ub^{S57C}

Table S3: Shift list comparison of ^1H - ^{15}N HSQC spectra of Tb- and Lu-DOTA-M8-(4R4S)-Ub^{S57C}.

Residue 2	Reson. 1	Shift Tb	Shift Lu	Reson. 2	Shift Tb	Shift Lu
2Gln	H	9.056	8.893	N	122.851	122.677
3Ile	H	8.336	8.207	N	115.131	114.942
4Phe	H	8.755	8.542	N	118.768	118.565
5Val	H	9.323	9.220	N	121.324	121.196
6Lys	H	8.950	8.861	N	128.038	127.906
7Thr	H	8.733	8.683	N	115.644	115.563
12Thr	H	8.648	8.588	N	120.785	120.685
13Ile	H	9.557	9.483	N	127.815	127.745
14Thr	H	8.749	8.688	N	121.784	121.702
15Leu	H	8.769	8.680	N	125.318	125.215
16Glu	H	8.041	8.051	N	122.490	122.456
23Ile	H	8.637	8.424	N	121.410	121.172
25Asn	H	7.975	7.871	N	121.561	121.453
26Val	H	8.126	8.030	N	122.255	122.156
27Lys	H	8.550	8.487	N	119.070	118.974
28Ala	H	7.942	7.907	N	123.631	123.569
29Lys	H	7.803	7.783	N	120.311	120.267
30Ile	H	8.244	8.217	N	121.479	121.413
31Gln	H	8.499	8.481	N	123.676	123.595
32Asp	H	7.946	7.942	N	119.767	119.729
33Lys	H	7.337	7.333	N	115.471	115.426
34Glu	H	8.668	8.659	N	114.432	114.377
39Asp	H	8.475	8.461	N	113.735	113.690
40Gln	H	7.762	7.757	N	116.912	116.882
41Gln	H	7.438	7.425	N	118.232	118.183
42Arg	H	8.437	8.425	N	123.197	123.153
43Leu	H	8.790	8.770	N	124.528	124.470
44Ile	H	9.068	9.031	N	122.286	122.229
45Phe	H	8.746	8.772	N	124.961	125.013
48Lys	H	7.794	7.897	N	122.041	122.120
49Gln	H	8.493	8.585	N	123.127	123.187
66Thr	H	8.922	8.654	N	117.706	117.424
67Leu	H	9.484	9.341	N	127.881	127.741
68His	H	9.204	9.141	N	119.526	119.394
69Leu	H	8.234	8.189	N	123.916	123.860
70Val	H	9.127	9.110	N	126.555	126.511
71Leu	H	8.053	8.048	N	123.016	122.985

Shift list comparison of ^1H - ^{15}N HSQC spectra of Dy- and Lu-DOTA-M8-(4R4S)-SSPy attached to selectively ^{15}N leucine labeled human carbonic anhydrase II S50C

Table S4: Shift list comparison of ^1H - ^{15}N HSQC spectra of Dy- and Lu-DOTA-M8-(4R4S)-SSPy attached to selectively ^{15}N leucine labeled human carbonic anhydrase II S50C.

Residue	Reson. 1	Shift Dy	Shift Lu	Reson. 2	Shift Dy	Shift Lu
44Leu	H	7.428	7.026	N	123.577	123.075
57Leu	H	8.903	8.953	N	117.617	117.504
60Leu	H	8.606	8.724	N	123.687	123.807
84Leu	H	7.027	7.818	N	119.594	120.387
100Leu	H	7.330	7.367	N	119.019	118.991
118Leu	H	9.768	9.915	N	130.343	130.468
120Leu	H	8.529	9.056	N	123.272	123.769
141Leu	H	7.898	8.790	N	114.606	115.551
144Leu	H	7.711	8.503	N	127.729	128.402
148Leu	H	8.202	8.253	N	120.314	120.333
157Leu	H	7.755	7.580	N	116.958	116.756
164Leu	H	7.127	7.099	N	121.671	121.665
184Leu	H	7.597	7.258	N	117.646	117.255
185Leu	H	7.014	6.662	N	113.483	113.152
189Leu	H	8.967	8.779	N	125.949	125.924
198Leu	H	7.669	8.049	N	117.979	118.323
203Leu	H	8.883	9.153	N	120.054	120.218
204Leu	H	5.657	5.975	N	111.090	111.317
212Leu	H	8.807	9.004	N	126.459	126.480
224Leu	H	8.050	8.052	N	120.897	120.933
229Leu	H	7.089	7.127	N	119.289	119.280
240Leu	H	8.625	8.678	N	125.782	125.722
251Leu	H	8.727	8.721	N	125.895	125.851

Shift list comparison of ^1H - ^{15}N HSQC spectra of Tm- and Lu-DOTA-M8-(4R4S)-SSPy attached to selectively ^{15}N leucine labeled human carbonic anhydrase II S50C

Table S5: Shift list comparison of ^1H - ^{15}N HSQC spectra of Tm- and Lu-DOTA-M8-(4R4S)-SSPy attached to selectively ^{15}N leucine labeled human carbonic anhydrase II S50C.

Residue	Reson. 1	Shift Tm	Shift Lu	Reson. 2	Shift Tm	Shift Lu
44Leu	H	6.706	7.026	N	122.731	123.075
57Leu	H	8.932	8.953	N	117.441	117.504
60Leu	H	8.834	8.724	N	123.846	123.807
84Leu	H	10.247	7.818	N	122.693	120.387
100Leu	H	7.399	7.367	N	119.020	118.991
118Leu	H	10.103	9.915	N	130.679	130.468
120Leu	H	9.903	9.056	N	124.572	123.769
141Leu	H	10.443	8.790	N	117.498	115.551
144Leu	H	9.892	8.503	N	129.784	128.402
148Leu	H	8.293	8.253	N	120.454	120.333
157Leu	H	7.169	7.580	N	116.390	116.756
164Leu	H	6.975	7.099	N	121.642	121.665
184Leu	H	6.490	7.258	N	116.611	117.255
185Leu	H	5.846	6.662	N	112.569	113.152
189Leu	H	8.081	8.779	N	125.282	125.924
198Leu	H	8.691	8.049	N	119.004	118.323
203Leu	H	9.617	9.153	N	120.734	120.218
204Leu	H	6.523	5.975	N	111.853	111.317
212Leu	H	9.271	9.004	N	126.839	126.480
224Leu	H	7.975	8.052	N	120.950	120.933
229Leu	H	7.136	7.127	N	119.328	119.280
240Leu	H	8.805	8.678	N	126.027	125.722
251Leu	H	8.887	8.721	N	126.041	125.851

Shift list comparison of ^1H - ^{15}N HSQC spectra of Tb- and Lu-DOTA-M8-(4R4S)-SSPy attached to selectively ^{15}N leucine labeled human carbonic anhydrase II S50C

Table S6: Shift list comparison of ^1H - ^{15}N HSQC spectra of Tb- and Lu-DOTA-M8-(4R4S)-SSPy attached to selectively ^{15}N leucine labeled human carbonic anhydrase II S50C.

Residue	Reson. 1	Shift Tb	Shift Lu	Reson. 2	Shift Tb	Shift Lu
44Leu	H	7.139	7.026	N	123.310	123.075
57Leu	H	8.904	8.953	N	117.594	117.504
60Leu	H	8.572	8.724	N	123.592	123.807
84Leu	H	6.931	7.818	N	119.584	120.387
100Leu	H	7.252	7.367	N	119.008	118.991
118Leu	H	9.749	9.915	N	130.328	130.468
120Leu	H	8.603	9.056	N	123.364	123.769
141Leu	H	7.967	8.790	N	114.720	115.551
144Leu	H	7.836	8.503	N	127.885	128.402
148Leu	H	8.151	8.253	N	120.320	120.333
157Leu	H	7.688	7.580	N	116.927	116.756
164Leu	H	7.046	7.099	N	121.742	121.665
184Leu	H	7.559	7.258	N	117.626	117.255
185Leu	H	6.951	6.662	N	113.460	113.152
189Leu	H	8.743	8.779	N	125.882	125.924
198Leu	H	7.631	8.049	N	117.988	118.323
203Leu	H	8.865	9.153	N	120.207	120.218
212Leu	H	8.760	9.004	N	126.361	126.480
224Leu	H	7.987	8.052	N	120.869	120.933
229Leu	H	7.016	7.127	N	119.238	119.280
240Leu	H	8.566	8.678	N	125.810	125.722
251Leu	H	8.676	8.721	N	125.874	125.851

Density functional theory calculations of lanthanide complexes – stabilization energies

Table S7: Stabilization energies obtained with or without coordinated water molecule in vacuo and water solvent (L = DOTA-M8-(4R4S)-SSPy).

Calculated energies of the different complexes M(L) (E_h / Hartree)		Stabilization (kJ/mol)
Lu(L) TSAP -16713.85773	Lu(L) SAP -16713.87043	-33.3
Lu(L) TSAP solv. -16713.95067	Lu(L) SAP solv. -16713.96020	-25.0
Lu(L)·H ₂ O TSAP -16790.34318	Lu(L)·H ₂ O SAP -16790.36243	-50.5
Lu(L)·H ₂ O TSAP solv. -16790.42518	Lu(L)·H ₂ O SAP solv. -*	-
Tm(L) TSAP -15802.89887	Tm(L) SAP -15802.90695	-21.2
Tm(L) TSAP solv. -15802.98688	Tm(L) SAP solv. -15802.99067	-10.0
Tm(L)·H ₂ O TSAP -15879.38336	Tm(L)·H ₂ O SAP -15879.39700	-35.8
Tm(L)·H ₂ O TSAP solv. -*	Tm(L)·H ₂ O SAP solv. -*	-
Dy(L) TSAP -14504.81886	Dy(L) SAP -14504.82753	-22.8
Dy(L) TSAP solv. -14504.90612	Dy(L) SAP solv. -14504.91021	-10.7
Dy(L)·H ₂ O TSAP -14581.30633	Dy(L)·H ₂ O SAP -14581.31911	-33.6
Dy(L)·H ₂ O TSAP solv. -14581.38820	Dy(L)·H ₂ O SAP solv. -14581.39086	-7.0

-*: No converged, meaningful, nine-coordinated complex obtained in the performed DFT calculations.

^1H spectrum of Dy-DOTA-M8-(4R4S)-SSPy

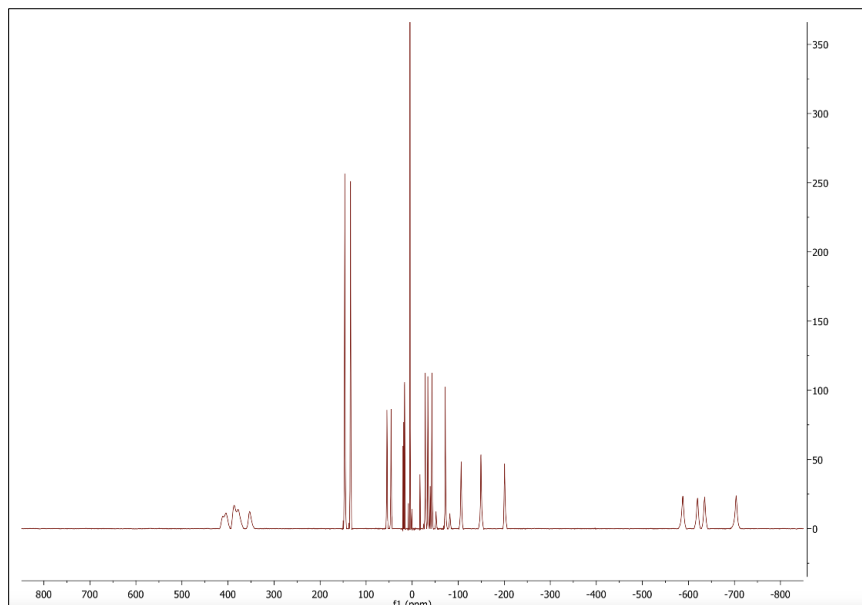


Figure S8: ^1H spectrum of Dy-DOTA-M8-(4R4S)-SSPy.

^1H spectrum of Tm-DOTA-M8-(4R4S)-SSPy

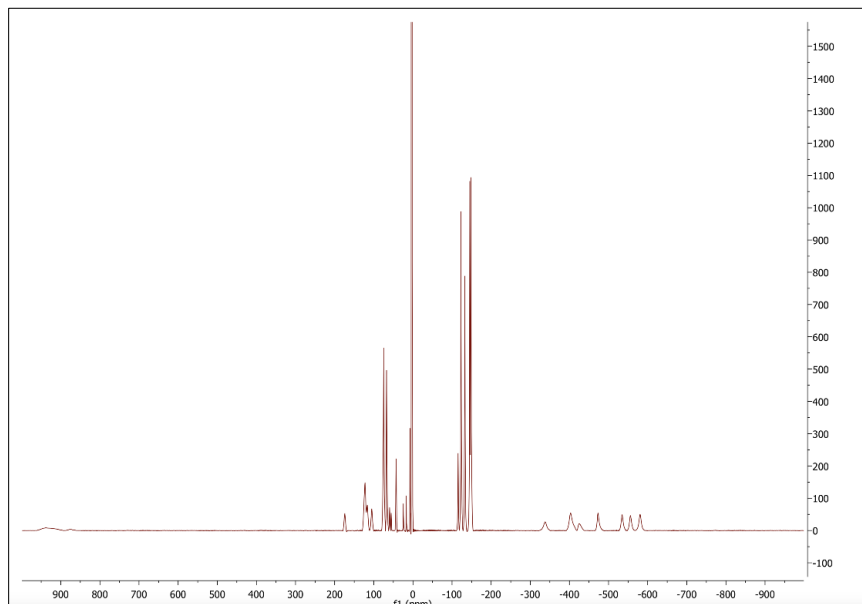


Figure S9: ^1H spectrum of Tm-DOTA-M8-(4R4S)-SSPy.

¹H spectrum of Tb-DOTA-M8-(4R4S)-SSPy

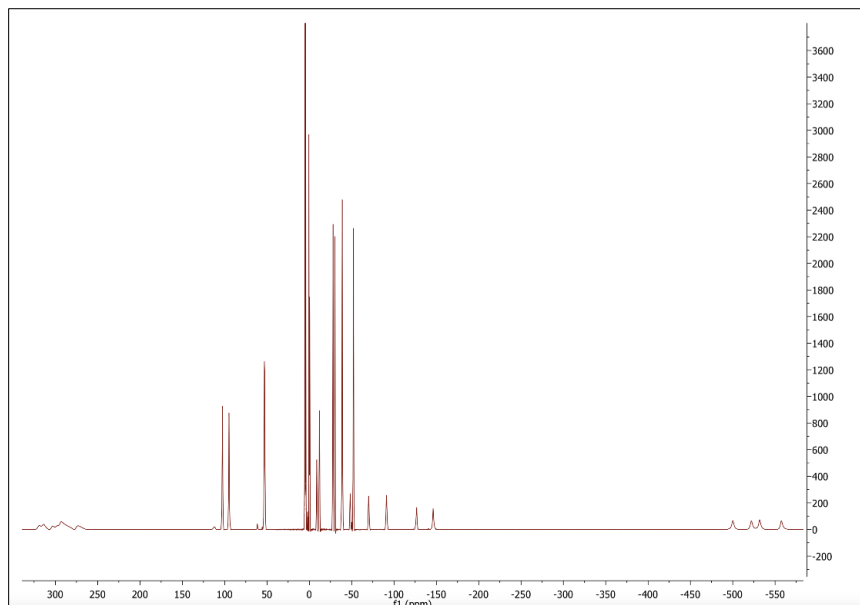


Figure S10: ¹H spectrum of Tb-DOTA-M8-(4R4S)-SSPy.

¹H spectrum of Yb-DOTA-M8-(4R4S)-SSPy

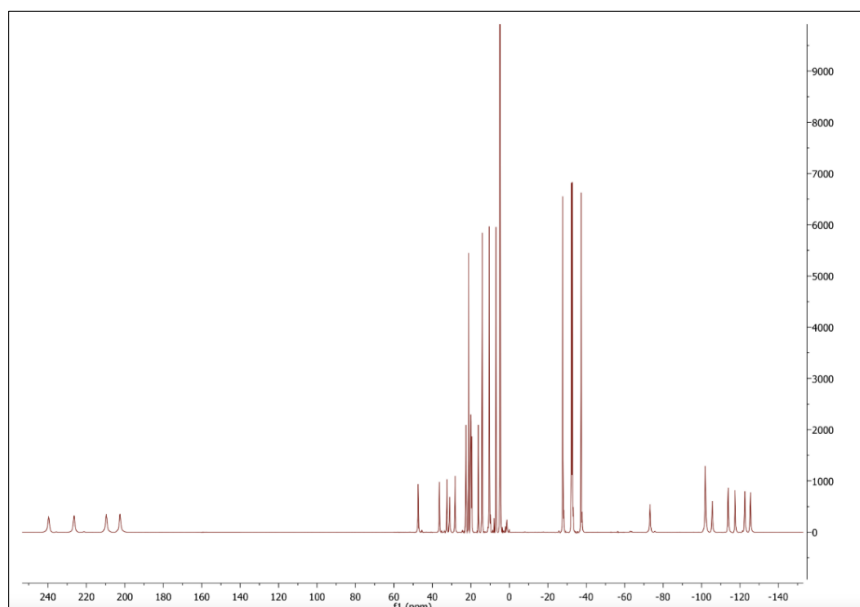
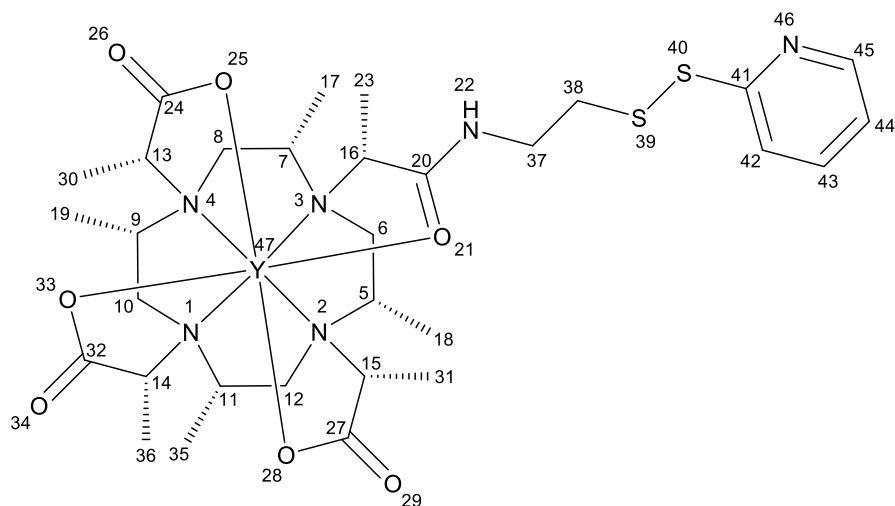


Figure S11: ¹H spectrum of Yb-DOTA-M8-(4R4S)-SSPy.

NMR Assignment of Y-DOTA-M8-(4R4S)-SSPy



¹H NMR (600 MHz, Deuterium Oxide) δ 8.60 (d, $J = 5.1$ Hz, 1H, H45), 8.41 – 8.34 (m, 1H, H43), 8.24 (d, $J = 8.4$ Hz, 1H, H42), 7.79 – 7.72 (m, 1H, H44), 4.00 (q, $J = 7.3$ Hz, 1H, H16), 3.79 – 3.66 (m, 3H, H13,14,37), 3.60 (q, $J = 7.3$ Hz, 1H, H15), 3.50 (ddd, $J = 14.2, 5.8$ Hz, 1H, H37), 3.12 – 2.94 (m, 9H, H6_{ax},7,8_{ax},9,10_{ax},11,12_{ax},38), 2.94 – 2.86 (m, 1H, H5), 2.61 – 2.53 (m, 4H, H6_{eq},8_{eq},10_{eq},12_{eq}), 1.53 (d, $J = 7.2$ Hz, 3H, H23), 1.42 (d, $J = 7.2$ Hz, 3H, H30), 1.39 (d, $J = 7.2$ Hz, 3H, H36), 1.38 (d, $J = 7.6$ Hz, 3H, H31), 1.14 – 1.07 (m, 12H, H17,18,19,35).

¹³C NMR (151 MHz, Deuterium Oxide, HMQC/HMBC, 0.01% TSP) δ 186.0 (C32), 185.9 (C24), 185.7 (C27), 182.9 (C20), 158.7 (C41), 148.3 (C43), 145.7 (C45), 128.2 (C42), 126.8 (C44), 69.8 (C15), 69.3 (C14), 69.3 (C13), 67.3 (C16), 63.6 (C5), 63.2 (C9), 63.2 (C7), 63.1 (C11), 49.0 (C6), 49.0 (C8), 49.0 (C10), 49.0 (C12), 42.0 (C37), 40.2 (C38), 16.1 (C17), 16.1 (C18), 16.1 (C19), 16.1 (C23), 16.1 (C31), 16.1 (35), 16.0 (C36), 15.9 (C30).

¹H spectrum of Y-DOTA-M8-(4R4S)-SSPy

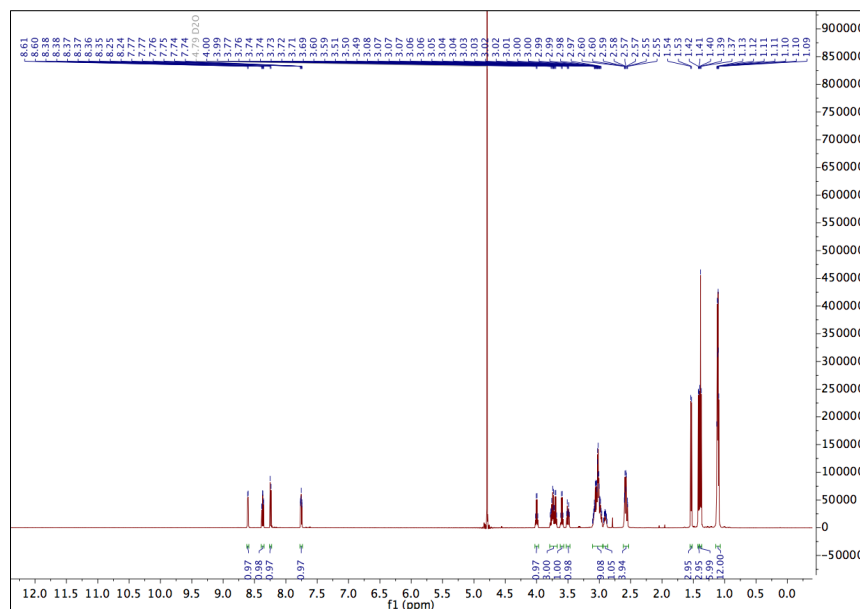


Figure S12: ¹H spectrum of Y-DOTA-M8-(4R4S)-SSPy.

ROESY spectrum of Y-DOTA-M8-(4R4S)-SSPy

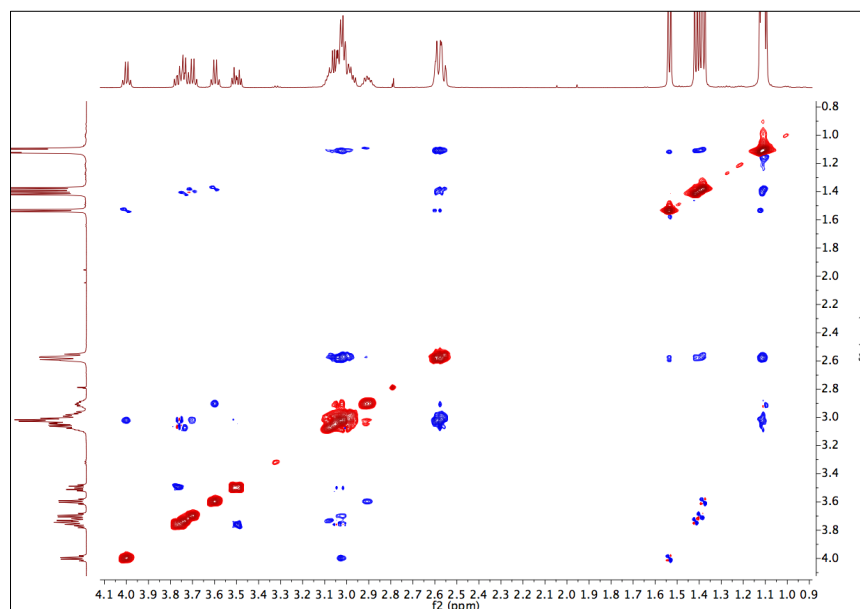


Figure S13: ROESY spectrum of Y-DOTA-M8-(4R4S)-SSPy.

COSY spectrum of Y-DOTA-M8-(4R4S)-SSPy

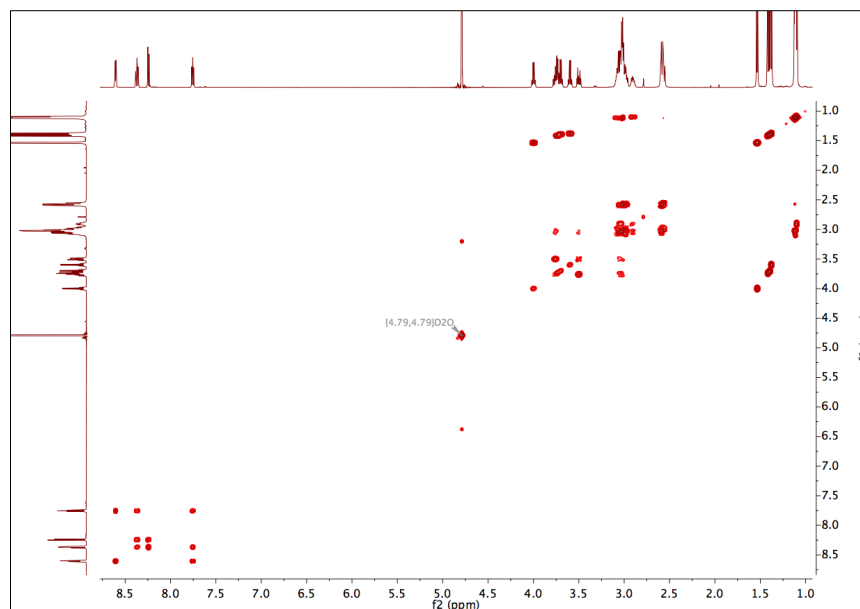


Figure S14: COSY spectrum of Y-DOTA-M8-(4R4S)-SSPy.

HMQC spectrum of Y-DOTA-M8-(4R4S)-SSPy

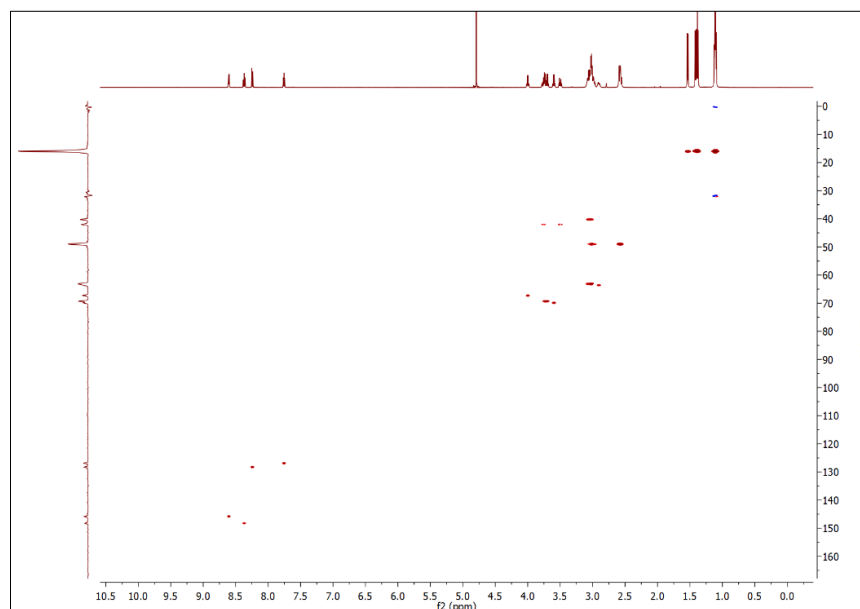


Figure S15: HMQC spectrum of Y-DOTA-M8-(4R4S)-SSPy.

HMBC spectrum of Y-DOTA-M8-(4R4S)-SSPy

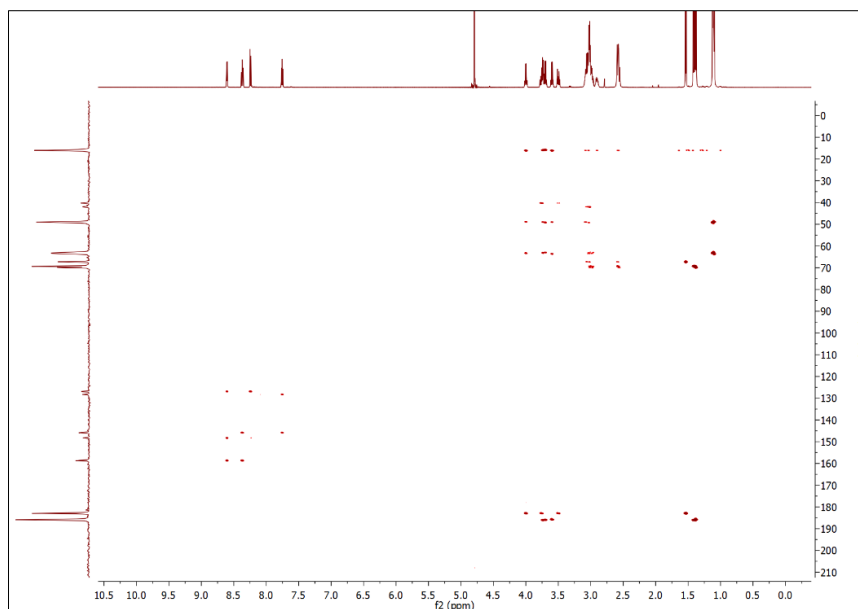


Figure S16: HMBC spectrum of Y-DOTA-M8-(4R4S)-SSPy.

3.2 A sterically overcrowded, isopropyl-substituted lanthanide chelating tag for protein pseudocontact shift NMR spectroscopy: Synthesis of its macrocyclic scaffold and benchmarking on ubiquitin S57C and hCA II S166C

Protein NMR

A Sterically Overcrowded, Isopropyl-Substituted, Lanthanide-Chelating Tag for Protein Pseudocontact Shift NMR Spectroscopy: Synthesis of its Macrocyclic Scaffold and Benchmarking on Ubiquitin S57C and hCA II S166C

Daniel Joss, Maria-Sophie Bertrams, and Daniel Häussinger*^[a]

Abstract: A sterically overcrowded lanthanide-chelating tag has been synthesized in order to investigate the influence on the obtained pseudocontact shifts and the anisotropic part of the magnetic susceptibility tensor compared to those of its predecessor DOTA-M8-(4*R*,4*S*)-SSPy. For the first time, a concise synthetic route is presented for isopropyl-substituted cyclen, the macrocyclic scaffold of the lanthanide-chelating tag, delivering the macrocycle in an overall yield of 6% over 11 steps. The geometry of the lutetium

complex has been assigned by ROESY experiments, adopting exclusively a $\Lambda(\delta\delta\delta\delta)$ conformation, and DFT calculations have confirmed a stabilization of 32.6 kJ mol^{-1} compared to the $\Delta(\delta\delta\delta\delta)$ conformer. The highly rigidified lanthanide-chelating tag induces strong pseudocontact shifts of up to 6.5 ppm on ubiquitin S57C, shows significantly improved tensor properties compared to those of its predecessor, and constitutes a highly promising starting point for the further development of lanthanide-chelating tags.

Introduction

Pseudocontact shifts (PCS) and residual dipolar couplings (RDC) generated by lanthanide-chelating tags (LCT) yield valuable structural restraints for the analysis of structure, dynamics, and ligand-binding of proteins in solution.^[1–20] In order to access valuable structural restraints by PCS measurements, stereospecifically methyl-substituted 1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid (DOTA)-based chelators present a suitable scaffold.^[9,19–22] As investigated by Ranganathan et al., the methyl substituents on the nitrogen-containing macrocycle in a lanthanide complex adopt an equatorial-upper position,^[23,24] that is, the substituents point away from the metal centre in order to accommodate the metal ion in the ligand cavity. In order to obtain only one signal set in ^1H - ^{15}N HSQC experiments and to induce strong paramagnetic effects for the investigation of biomacromolecules in solution, it is mandatory to lock not only the 12-membered ring,^[9] but also the pendant arms of the chelator, in one conformation. As shown by Joss et al., Tm- and Dy-DOTA-M8-(4*R*,4*S*)-SSPy adopt a $\Lambda(\delta\delta\delta\delta)$ conformation, resulting in a single set of peaks in ^1H - ^{15}N HSQC spectra.^[19]

We surmised that introducing isopropyl substituents on the macrocyclic scaffold would yield a lanthanide-chelating tag with improved tensor properties compared to those of the DOTA-M8-(4*R*,4*S*)-SSPy tag, since translational and rotational movements of the tag relative to the protein should be significantly hindered. Furthermore, any remaining small-amplitude vibrations of the pendant arms should be decreased by the sterically more compact packing of the isopropyl groups. We therefore envisaged the synthesis of a new isopropyl-substituted lanthanide-chelating tag (Figure 1) and its corresponding macrocyclic scaffold based on some previously published synthetic routes to related targets.^[8,9,21,22,25,26]

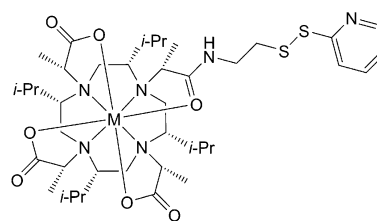


Figure 1. Structure of a metal complex of P4M4-DOTA in SAP conformation.

However, published synthetic routes for the tetra-methyl analogue did not yield useful quantities in the case of the tetra-isopropyl cyclen. Therefore, we developed a new synthetic route involving two reductive aminations as key steps to overcome the hampered reactivity due to highly sterically crowded substrates. Following the total synthesis of the macrocyclic scaffold, we produced Dy-, Tm-, and Lu-P4M4-DOTA (P4M4-

[a] D. Joss, M.-S. Bertrams, Dr. D. Häussinger
Department of Chemistry
University of Basel
St. Johans-Ring 19
4056 Basel (Switzerland)
E-mail: daniel.haeussinger@unibas.ch

Supporting information and the ORCID identification numbers for the authors of this article can be found under:
<https://doi.org/10.1002/chem.201901692>

DOTA: (2*R*,2'*R*,2''*R*)-2,2',2''-((2*S*,5*S*,8*S*,11*S*)-2,5,8,11-tetraisopropyl-10-((*R*)-1-oxo-1-((2-(pyridin-2-yl)disulfanyl)ethyl)amino)propan-2-yl)-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)tripropionate) and conjugated each lanthanide complex to ubiquitin S57C and a selectively ¹⁵N-leucine-labelled human carbonic anhydrase II (hCA II) S166C protein construct. The obtained spectra, tensor sizes, and orientations, as well as tag mobilities of the complexes when attached to the protein, were analysed.

Results and Discussion

Synthesis of the macrocyclic scaffold

In order to gain synthetic access to the macrocyclic scaffold of the P4M4 tag, we attempted its synthesis based on published routes to related targets.^[21,22,25,26] However, all previously used synthetic routes to related targets failed in the case of our isopropyl-substituted macrocycle or gave only mixtures of products not useful for the generation of a lanthanide-chelating tag suitable for PCS NMR spectroscopy. The stepwise aziridine synthesis published in 2009 by Kamioka et al. constitutes an intellectually very appealing approach, but the reagents employed are known to be very toxic and highly carcinogenic.^[8] We therefore decided to develop a new route towards this macrocyclic scaffold involving formation of the highly sterically crowded tetra-valine analogue by using reductive aminations as key reactions (Figure 2). For the high-yielding macrocyclization, we referred to previous knowledge accumulated by our group.^[20]

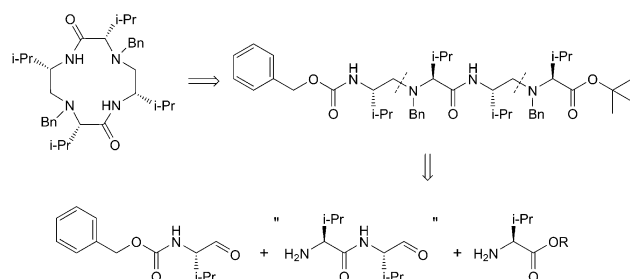


Figure 2. Key retrosynthetic steps of the synthesis of the new tetra-isopropyl-substituted macrocyclic polyamine.

Reductive aminations were envisaged as ideal reactions to couple the different parts of the sterically overcrowded tetra-valine analogue, since a synthetic strategy based on a (7-azabenzotriazol-1-yl)tetramethyluronium hexafluorophosphate (HATU) coupling of two valine-derived dimers did not yield the desired product in sufficient quantity.

The final synthetic route to the new macrocycle comprises 11 steps and is depicted in Figures 3, 4, and 5.

Cbz-protected valine (**1**) and valine methyl ester (**2**) were used as inexpensive and readily available chiral building blocks. They were coupled employing 1-propanephosphonic acid anhydride (T3P) to give dipeptide **3**. Subsequent reduction to alcohol **4** and mild oxidation with 2-iodoxybenzoic acid

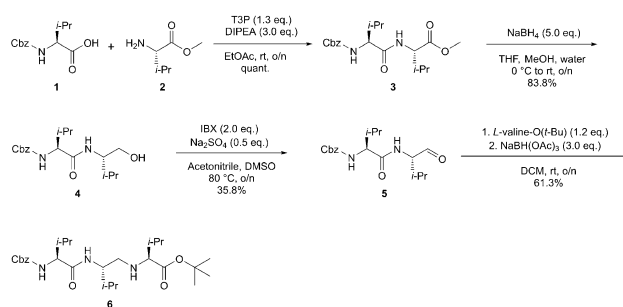


Figure 3. Formation of the required sterically overcrowded tetra-valine analogue (Part I).

(IBX)^[27,28] furnished the corresponding aldehyde **5**, which was then subjected to reductive amination with valine *tert*-butyl ester to generate trimeric structure **6** (Figure 3).

Cbz deprotection of the trimeric peptide analogue **6** was accomplished by hydrogenation, and the product was then reacted with Cbz-valinal in a reductive amination to give tetrameric peptide analogue **8**. In order to obtain the sterically overcrowded but flexible tetra-valine analogue **9**, the tetrameric structure **8** was benzylated with benzyl bromide in the presence of potassium iodide as catalyst, accelerating the reaction through an in situ Finkelstein conversion of benzyl bromide to benzyl iodide (Figure 4).

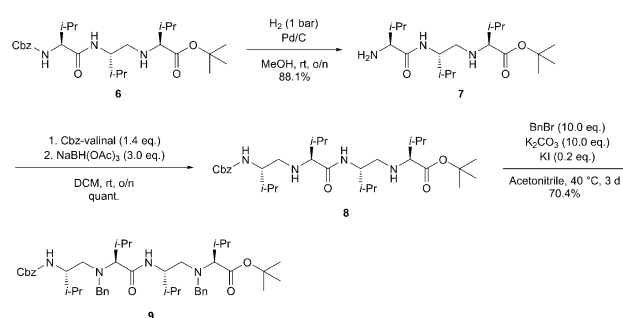


Figure 4. Formation of the required sterically overcrowded tetra-valine analogue (Part II).

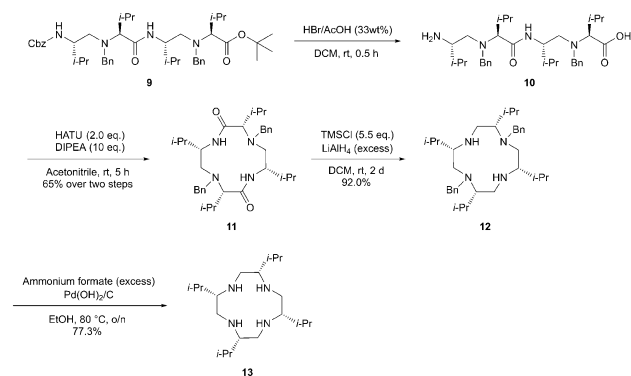


Figure 5. Cyclization of the flexible tetra-valine analogue and subsequent steps towards the final product.

Tetraline analogue **9** was deprotected using HBr in acetic acid and readily underwent cyclization to give bislactam **11**. Bislactam **11** was then transformed in two steps, a TMSCl-activated lithium aluminium hydride reduction^[29] and a benzyl deprotection by transfer hydrogenation, to the final product (**13**) (Figure 5).

The crystal structure obtained for the final product shows the isopropyl substituents in four-fold equatorial positions (Figure 6). While the pro-(*R*) methyl group of the isopropyl substituent is arranged in a plus-synclinal conformation, the pro-(*S*) methyl group is positioned in an antiperiplanar fashion (dihedral angle $N-C_{CH}-C_{CH}-C_{Me}$).

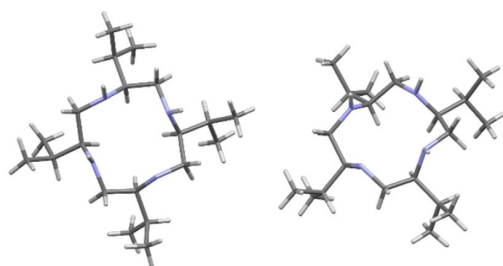


Figure 6. Crystal structure of the isopropyl-substituted cyclen (CCDC-1917602): top view (left), diagonal view confirming the all-(*S*) configuration and the preferred conformation with all isopropyl substituents in equatorial positions (right).

Synthesis of Ln-P4M4-DOTA

Ln-P4M4-DOTA was synthesized according to a similar reaction sequence as in the procedure reported by Häussinger et al. for the Lu-DOTA-M8-(4*R*,4*S*)-SSPy tag.^[9] For the alkylation steps, solvents were modified and excess triflate was used. The final tag showed only one Lu-P4M4-DOTA species, which was assigned by ROESY experiments as adopting a $\Delta(\delta\delta\delta\delta)$ (SAP) conformation (spectra in the Supporting Information, p. 22). The two strongly paramagnetic complexes most likely adopt the same conformation based on the virtually identical retention time in HPLC measurements (HPLC traces in the Supporting Information, Figures S1–S3) and the energetic bias of the LCT towards an SAP conformation of 26.6–32.6 kJ mol⁻¹ estimated by DFT calculations using an implicit water solvent model (energies in Table S10).

Properties of the isopropyl-substituted lanthanide-chelating tag when attached to ubiquitin S57C

In order to test the properties of the new lanthanide-chelating tag, we attached it to a ubiquitin S57C construct and subjected it to HSQC experiments. Despite the increased hydrophobicity of Ln-P4M4-DOTA compared to that of its methyl-substi-

tuted predecessor (Ln-DOTA-M8-(4*R*,4*S*)),^[19] Ln-P4M4-DOTA is freely water-soluble and caused no precipitation of the investigated proteins upon addition and conjugation of the LCT. In a first step, we compared untagged protein to diamagnetic Lu-P4M4-DOTA-Ub^{S57C} and found only minor, very localized shifts in the immediate vicinity of the tagging site (Figure S24). The tag showed PCS of up to 6.5 ppm when attached to ubiquitin S57C, and the PCS and tensor anisotropies observed for Ln-P4M4-DOTA upon attachment to the protein were significantly increased compared to those of its methylated predecessor (Figures 7 and 8 and 9; Table 1). The magnitude of the obtained PCS and anisotropy parameters as well as the straightforward application places Ln-P4M4-DOTA among the best performing LCTs currently available.^[3,9,11,17,19,30,31]

The enhanced PCS could be attributed to one or more of the following factors: (i) decreased small-amplitude vibrations of the scaffold, (ii) different positioning of the oxygen donor atoms of the side arms relative to the nitrogen donors of the basal cyclen, which was shown by Mironov et al.^[32] to significantly affect the observed anisotropies, (iii) altered mobility or rotation of the tag on the protein surface. As can be seen from the calculated structure of the Lu-P4M4-DOTA complex, the isopropyl groups on the backbone lead to extreme overcrowding of the chelating cage (Figure 10).

ROESY experiments support this view and show significantly stronger ROE correlations from the side-arm methyl groups and CH groups of the basic ring scaffold to the pro-(*R*) methyl substituents of the isopropyl groups, indicating only one preferred conformation (Figure 10).

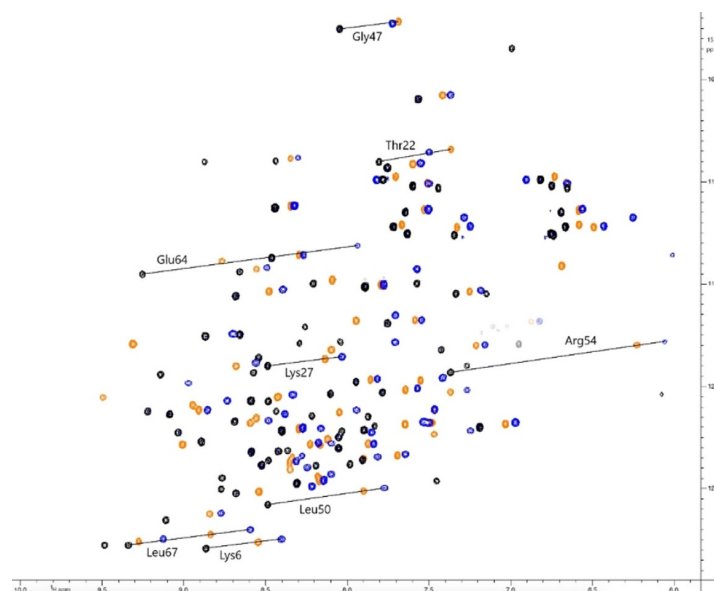


Figure 7. Overlay of ¹H-¹⁵N HSQC spectra of Tm-DOTA-M8-(4*R*,4*S*)-Ub^{S57C} (pos. contours: orange, neg. contours: pale-orange) and Tm- (pos. contours: blue, neg. contours: pale-blue) and Lu-P4M4-DOTA-Ub^{S57C} (pos. contours: black, neg. contours: grey) measured in 10 mM phosphate buffer at pH 6.0 and 298 K on a 600 MHz Bruker Avance III NMR spectrometer equipped with a cryogenic QCI-F probe. Annotated cross-peaks highlight large PCS and the increase in PCS between Tm-P4M4-DOTA and Tm-DOTA-M8-(4*R*,4*S*)-SSPy. Negative peaks with a ¹H shift around 7.0 ppm and a ¹⁵N shift around 117.5 ppm constitute aliased signals from arginine side chains.

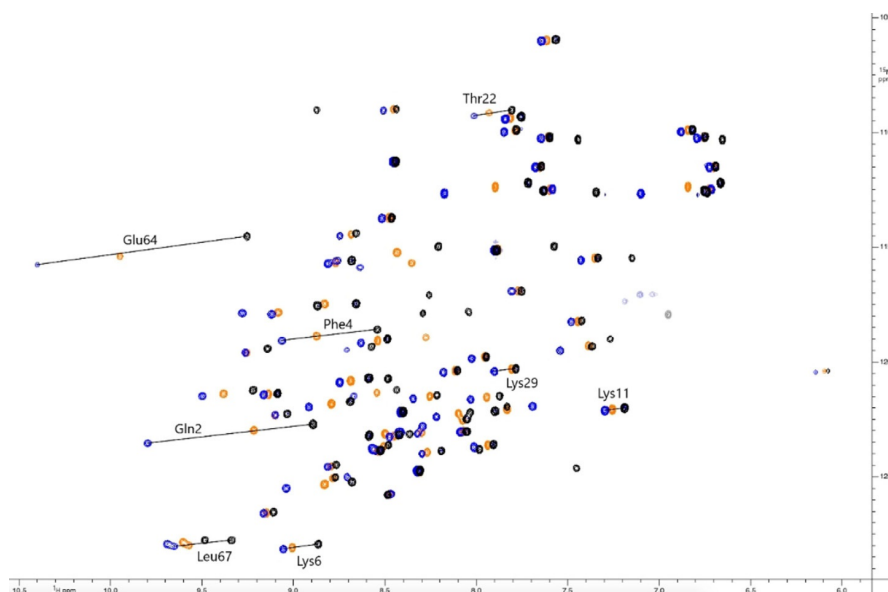


Figure 8. Overlay of ^1H - ^{15}N HSQC spectra of Dy-DOTA-M8-(4R,4S)-Ub^{557C} (pos. contours: orange, neg. contours: pale-orange) and Dy- (pos. contours: blue, neg. contours: pale-blue) and Lu-P4M4-DOTA-Ub^{557C} (pos. contours: black, neg. contours: grey) measured in 10 mM phosphate buffer at pH 6.0 and 298 K on a 600 MHz Bruker Avance III NMR spectrometer equipped with a cryogenic QCI-F probe. Annotated cross-peaks highlight large PCS and the increase in PCS between Dy-P4M4-DOTA and Dy-DOTA-M8-(4R,4S)-SSPy. Negative peaks with a ^1H shift around 7.0 ppm and a ^{15}N shift around 117.5 ppm constitute aliased signals from arginine side chains.

Table 1. Properties of the induced anisotropy of the magnetic susceptibility tensors of Ln-P4M4-DOTA measured on ubiquitin S57C at 298 K and pH 6.0.

LCT	No PCS*/ RDC**	Ln ³⁺	$\Delta\chi_{ax}$ [10 ⁻³² m ³]	$\Delta\chi_{rh}$ [10 ⁻³² m ³]	X_{metal} [Å]	Y_{metal} [Å]	Z_{metal} [Å]	α [°]	β [°]	γ [°]	Q [%]	Ref.
P4M4-DOTA	72*	Dy	-16.7	-7.5	16.4	13.6	7.1	149.7	43.0	42.7	5.2	This work
	70*	Tm	26.0	1.0	17.3	13.5	9.6	161.3	53.1	50.7	1.4	
DOTA-M8-(4R,4S)	74*	Dy	-8.7	-4.1	16.8	14.6	9.0	160.0	37.6	37.2	13.6	[19]
	68*	Tm	19.6	3.0	16.8	14.6	9.0	174.7	44.9	83.9	6.2	
P4M4-DOTA	72**	Dy	14.1	-4.8	16.4	13.6	7.1	-106.5	-131.8	158.0	31.4	This work
	70**	Tm	20.0	7.3	17.3	13.5	9.6	-115.2	-126.3	160.6	30.1	

In terms of the coordination geometry of the ligand, our DFT calculations showed that the position of the donor atoms differs by no more than 0.11 Å between the isopropyl and methyl variants of the LCT (Figures S25 and S26). These small

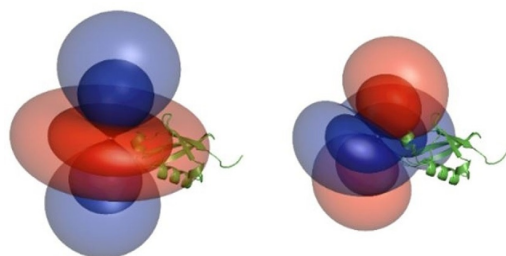


Figure 9. The tensors generated by the thulium complex (left) and dysprosium complex (right) and their relative orientation to ubiquitin S57C (PCS iso-surfaces set to 0.2 ppm (outer layer) and 0.8 ppm (inner layer)).

changes in the geometry of the LCT are also reflected in the intrinsic proton shift range of Dy-P4M4-DOTA, which is broadened by 70 ppm compared to that of its predecessor, and the intrinsic shift range of Tm-P4M4-DOTA, which is contracted by 50 ppm (spectra in the Supporting Information, p. 23).

In order to investigate the mobility of the newly developed lanthanide-chelating tag on the protein surface, the metal centre distribution was modelled using a method developed by Suturina et al. (Figure 11).^[33–35] In PCS NMR spectroscopy, the point paramagnetic centre approximation is usually applied. Since the protein is measured in solution at room temperature and LCTs maintain their flexibility, the metal position can be more realistically described in terms of its probability density. The approach developed by Suturina et al. allows extraction of the metal position probability distribution by a Tikhonov-regularized 3D reconstruction and a partial differential equation, which is used to describe the PCS on the basis of a non-point electron probability density.^[34, 36]

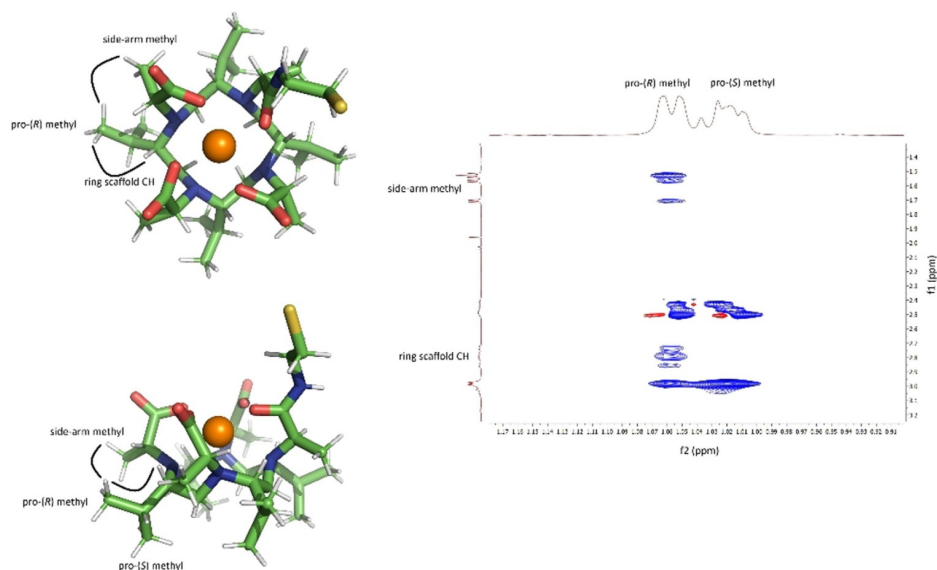


Figure 10. Left: Calculated DFT structure of Lu-P4M4-DOTA, top and side views, SPy activator omitted for clarity. Right: ROE correlations suggesting one preferred conformation of the isopropyl groups of Lu-P4M4-DOTA due to the steric overcrowding.

The results obtained for the calculated metal distribution showed that the tag was immobilized to the same extent for both P4M4-DOTA and DOTA-M8-(4*R*,4*S*)-SSPy (Figure 11).

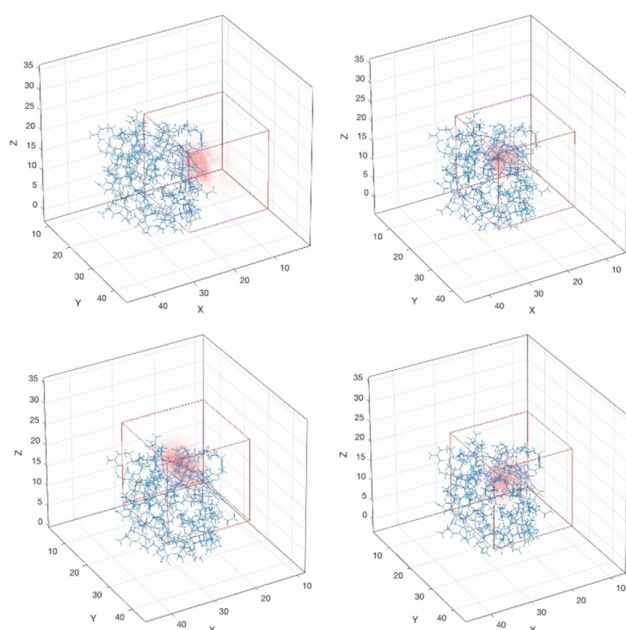


Figure 11. Distributions of the metal position for Tm-DOTA-M8-(4*R*,4*S*)-Ub^{557C} (upper left), Tm-P4M4-DOTA-Ub^{557C} (lower left), Dy-DOTA-M8-(4*R*,4*S*)-Ub^{557C} (upper right), and Dy-P4M4-DOTA-Ub^{557C} (lower right). The metal position distribution is depicted as a colour gradient (red = high occupancy, white = very low occupancy). The metal position is restricted to the boundaries of the red cube. The centre of this cube is given by the initial metal position obtained from the point paramagnetic centre approximation resulting from the fitting of the PCS using Spinach. The blue wireframe represents the ubiquitin structure (PDB 1UBI³⁷).

We can therefore conclude that the enhanced tensor properties of the P4M4-DOTA tag most probably arise principally from less averaging of the PCS by decreased rotation of the tag in its position on the surface of the protein. The decrease in the remaining small-amplitude vibrations of the donor atoms of the pendant arms of the ligand probably further enhances the size of the shifts obtained with the P4M4-DOTA tag.

Benchmarking on selectively ¹⁵N-leucine-labelled human carbonic anhydrase II S166C

In order to investigate the applicability of the newly developed LCT to larger proteins, we tagged selectively ¹⁵N-leucine-labelled hCA II S166C, a 29 kDa protein construct, with both the dysprosium and thulium complexes (Figures 12 and 13). Furthermore, we attached Tm- and Dy-DOTA-M8-(4*R*,4*S*)-SSPy to hCA II S166C in order to permit comparison of Ln-P4M4-DOTA and its predecessor.

The obtained anisotropy of the tensors on selectively ¹⁵N-leucine-labelled hCA II S166C (Table 2) was enhanced compared to that of the tensors for ubiquitin S57C (Table 1). Furthermore, the observed PCS and tensor anisotropies for P4M4-DOTA attached to hCA II S166C were significantly enhanced compared to those of the hCA II S166C-DOTA-M8-(4*R*,4*S*)-SSPy construct for both its dysprosium and thulium complexes. The anisotropy of the Dy tensor relative to the Tm tensor on the hCA II S166C protein construct remained similar to the ratio obtained for P4M4-DOTA-Ub^{557C} (Table 2, Figure 14).

The results obtained for ubiquitin S57C and hCA II S166C show that the Ln-P4M4-DOTA tag is applicable without restraints for PCS NMR spectroscopy of proteins as large as 29 kDa and shows promise to also be of value for larger pro-

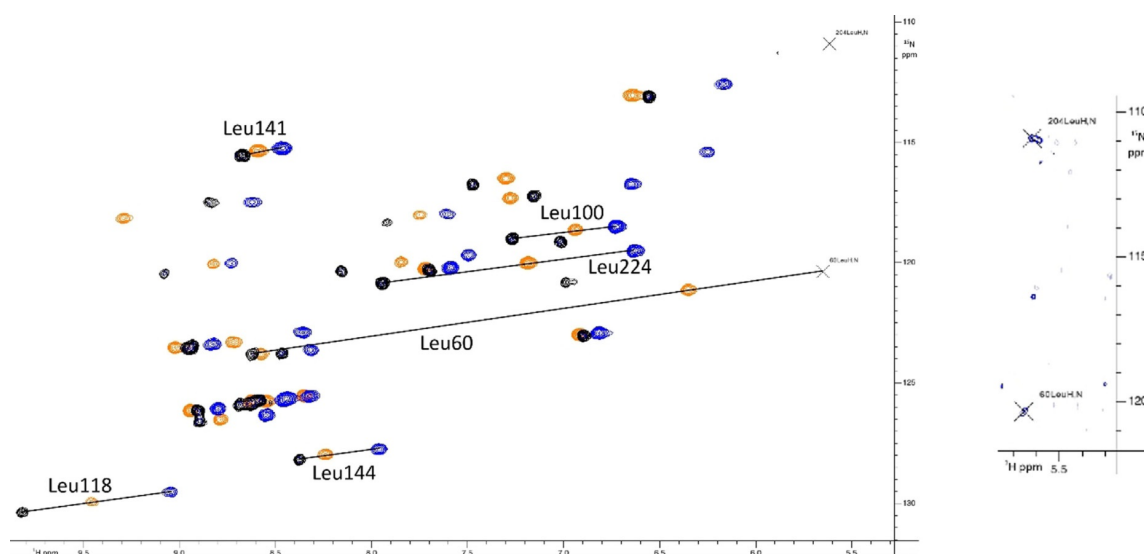


Figure 12. Overlay of ^1H - ^{15}N HSQC spectra of selectively ^{15}N -labelled hCA II S166C labelled with Tm-DOTA-M8-(4*R*,4*S*)-SSPy (pos. contours: orange, neg. contours: pale-orange) and Tm- (pos. contours: blue, neg. contours: pale-blue) and Lu-P4M4-DOTA (pos. contours: black, neg. contours: grey) measured in 10 mM phosphate buffer at pH 6.8 and 298 K on a 600 MHz Bruker Avance III NMR spectrometer equipped with a cryogenic QCI-F probe. Annotated cross-peaks highlight large PCS and the increase in PCS between Tm-P4M4-DOTA and Tm-DOTA-M8-(4*R*,4*S*)-SSPy. Cross-peaks close to the water resonance are shown in an extra panel to the right with lower contour levels.

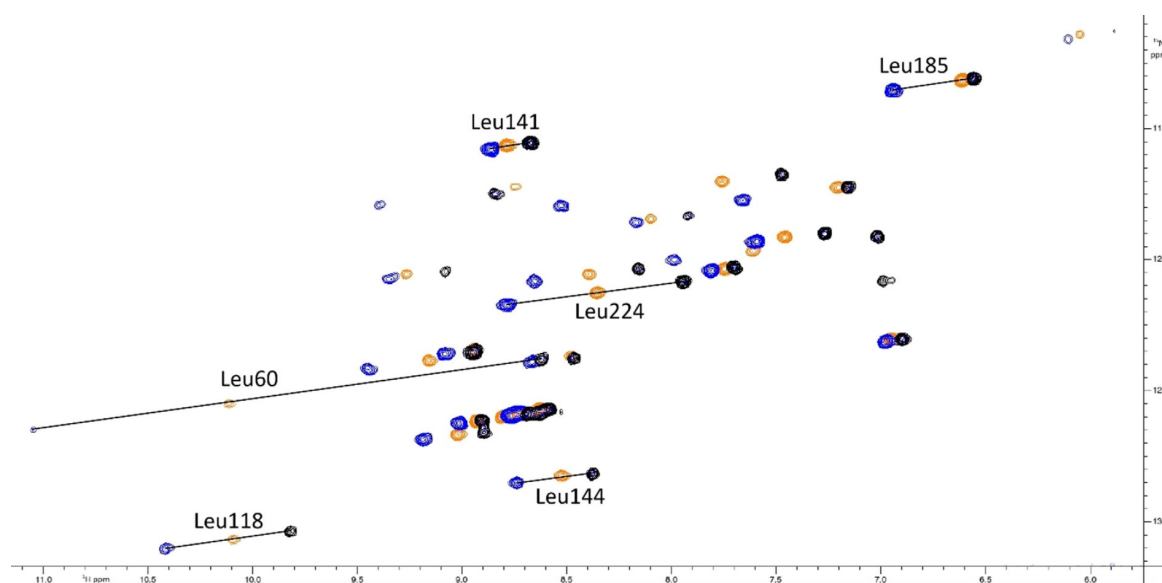


Figure 13. Overlay of ^1H - ^{15}N HSQC spectra of selectively ^{15}N -labelled hCA II S166C labelled with Dy-DOTA-M8-(4*R*,4*S*)-SSPy (pos. contours: orange, neg. contours: pale-orange) and Dy- (pos. contours: blue, neg. contours: pale-blue) and Lu-P4M4-DOTA (pos. contours: black, neg. contours: grey) measured in 10 mM phosphate buffer at pH 6.8 and 298 K on a 600 MHz Bruker Avance III NMR spectrometer equipped with a cryogenic QCI-F probe. Annotated cross-peaks highlight large PCS and the increase in PCS between Tm-P4M4-DOTA and Tm-DOTA-M8-(4*R*,4*S*)-SSPy.

teins, protein-protein complexes, and other biomacromolecules.

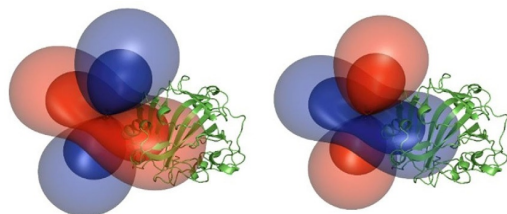
Conclusions

To conclude, two strongly paramagnetic lanthanide complexes of P4M4-DOTA are reported, both of which adopt a $\Lambda(\delta\delta\delta\delta)$ (SAP) conformation in solution and provide only one set of sig-

nals in HSQC experiments when attached to a protein. A synthetic route towards isopropyl-substituted cyclen is presented, which comprises 11 steps, gives an overall yield of 6%, and benefits from the use of readily available, inexpensive starting materials. The newly developed tag shows significantly enhanced anisotropy of the obtained tensors upon attachment to ubiquitin S57C and hCA II S166C compared to those for the DOTA-M8-(4*R*,4*S*)-SSPy analogue. Based on the mobility of the

Table 2. Properties of the induced paramagnetic susceptibility tensors measured on hCA II S166C at 298 K and pH 6.8.

LCT	No PCS*/ RDC**	Ln ³⁺	$\Delta\chi_{ax}$ [10 ⁻³² m ³]	$\Delta\chi_{rh}$ [10 ⁻³² m ³]	X_{metal} [Å]	Y_{metal} [Å]	Z_{metal} [Å]	α [°]	β [°]	γ [°]	Q [%]
P4M4-DOTA	44*	Dy	-29.7	-11.5	-15.5	-2.9	-11.4	144.1	76.8	93.6	3.2
	44*	Tm	41.9	17.1	-14.6	-2.3	-12.2	141.7	73.2	87.4	3.6
DOTA-M8-(4R,4S)	50*	Dy	-21.2	-4.5	-15.2	-3.2	-11.0	138.4	71.3	137.6	6.7
	50*	Tm	29.7	7.6	-14.5	-3.0	-10.8	142.3	66.5	117.6	5.8
P4M4-DOTA	44**	Tm	44.1	-11.1	-14.6	-2.3	-12.2	-156.3	-158.7	-10.4	46.0

**Figure 14.** The tensors generated by the thulium complex (left) and dysprosium complex (right) and their relative orientation to human carbonic anhydrase S166C (PCS isosurfaces set to 0.3 ppm (outer layer) and 1.3 ppm (inner layer)).

tag on the surface of ubiquitin S57C, the already very rigid basic ring scaffold of the predecessor, the similar position of the donor atoms, and the similar intrinsic shift ranges in proton spectra, it can be concluded that the improvement can mainly be attributed to less averaging of the PCS by a decreased rotation of the tag when attached to the protein. The tag has been shown to be applicable to proteins of 29 kDa in size and yields strong pseudocontact shifts of up to 6.5 ppm. In order to enable challenging applications in biomolecular NMR, as, for example, the structural investigation of extremely large proteins, protein complexes and other biopolymers, further development of high-performance lanthanide-chelating tags needs to be pursued.

Experimental Section

Synthesis

Lanthanide complexes were synthesized according to a similar reaction sequence as in the procedures reported by Häussinger et al. for the Lu-DOTA-M8-(4R,4S)-SSPy tag.^[9] For the macrocyclic scaffold, a new synthetic route was developed. The complete synthetic procedures as well as analytical data are available in the Supporting Information.

Expression of ubiquitin S57C and tagging reaction

Ubiquitin S57C was expressed as previously described by Sass et al.^[38] A sample of ubiquitin S57C (0.1 mM) was incubated with tris(2-carboxyethyl)phosphine (TCEP) (2 mM) at pH 7 overnight. The buffer was replaced by 10 mM phosphate containing 0.1 mM TCEP (pH 7.0 for ubiquitin S57C, pH 6.8 for hCA II S166C) by ultra-filtration (Amicon Ultra-4 Ultracel-3 K, cut-off 3 kDa). The remaining

minimal amount of TCEP was then removed by passage through a PD MiniTrap G-25 column. The obtained protein solution was added to a six-fold excess of Ln-P4M4-DOTA, the mixture was shaken at rt overnight, and completion of the reaction was confirmed by ESI-MS. The buffer was replaced by 10 mM phosphate (pH 6.0 for ubiquitin S57C, pH 6.8 for hCA II S166C) by ultra-filtration (Amicon Ultra-4 Ultracel-3 K, cut-off 3 kDa) and the sample was concentrated to 250 μ L for NMR experiments.

NMR measurements and determination of the paramagnetic susceptibility tensor

¹H-¹⁵N HSQC spectra were measured from solutions in 10 mM phosphate buffer at pH 6.0 (ubiquitin S57C) or pH 6.8 (hCA II S166C) at 298 K on a 600 MHz Bruker Avance III NMR spectrometer equipped with a cryogenic QCI-F probe. ¹H-¹⁵N HSQC in-phase/anti-phase (IPAP) spectra were measured from solutions in 10 mM phosphate buffer at pH 6.0 and 298 K on a 900 MHz Bruker Avance III NMR spectrometer equipped with a TXI probe (ubiquitin S57C) or in 10 mM phosphate buffer at pH 6.8 and 298 K on a 600 MHz Bruker Avance III NMR spectrometer equipped with a QCI-F probe (hCA II S166C). Assignments of the obtained NMR spectra were made using CcpNmr Analysis.^[33,39,40] The tensor properties were then obtained by fitting to the residues in secondary structure elements of ubiquitin (PDB 1UBI^[37]) or the leucine residues of hCA II (PDB 3KS3^[41]) using Numbat.^[42] Q-factors were calculated according to the standard equation described by Nitsche et al.^[1]

DFT calculations

DFT calculations were performed with the ORCA program package^[43] at the sciCORE facility of the University of Basel. For these calculations, BP86 was used as the functional,^[44,45] SARC-TZVP as the basis set for the ligands, and SARC2-QZVP as the basis set for the lanthanide metal. The calculations were performed using the relativistic ZORA approximation, and expedited by applying the RI approximation. To model the water solvent, a CPCM solvent model was implemented in the calculations.^[46]

Acknowledgements

The Chemistry Department of the University of Basel and the Swiss National Science Foundation grant 200021 130263 are acknowledged for financial support. Calculations were performed at sciCORE (<http://scicore.unibas.ch/>) scientific computing core facility at the University of Basel. Biological structures were generated using the open source software PyMOL (<http://www.pymol.org/>). We are indebted to A. Prescimone for

acquiring and solving the crystal structure. C. E. Housecroft, E. C. Constable, T. Müntener, and R. Vogel are acknowledged for helpful discussions.

Conflict of interest

The authors declare no conflict of interest.

Keywords: lanthanides · macrocycles · paramagnetic · protein nuclear magnetic resonance spectroscopy · pseudocontact shift

- [1] C. Nitsche, G. Otting, *Progr. Nucl. Magn. Res. Spectr.* **2017**, *98–99*, 20–49.
- [2] B.-B. Pan, F. Yang, Y. Ye, Q. Wu, C. Li, T. Huber, X.-C. Su, *Chem. Commun.* **2016**, *52*, 10237–10240.
- [3] T. Müntener, D. Häussinger, P. Selenko, F.-X. Theillet, *J. Phys. Chem. Lett.* **2016**, *7*, 2821–2825.
- [4] Y. Hikone, G. Hirai, M. Mishima, K. Inomata, T. Ikeya, S. Arai, M. Shirakawa, M. Sodeoka, Y. Ito, *J. Biomol. NMR* **2016**, *66*, 99–110.
- [5] W.-M. Liu, M. Overhand, M. Ubbink, *Coord. Chem. Rev.* **2014**, *273*, 2–12.
- [6] C. T. Loh, K. Ozawa, K. L. Tuck, N. Barlow, T. Huber, G. Otting, B. Graham, *Bioconjugate Chem.* **2013**, *24*, 260–268.
- [7] G. Otting, *Annu. Rev. Biophys.* **2010**, *39*, 387–405.
- [8] S. Kamioka, T. Takahashi, S. Kawauchi, H. Adachi, Y. Mori, K. Fujii, H. Uekusa, T. Doi, *Org. Lett.* **2009**, *11*, 2289–2292.
- [9] D. Häussinger, J.-R. Huang, S. Grzesiek, *J. Am. Chem. Soc.* **2009**, *131*, 14761–14767.
- [10] X.-C. Su, B. Man, S. Beeren, H. Liang, S. Simonsen, C. Schmitz, T. Huber, B. A. Messerle, G. Otting, *J. Am. Chem. Soc.* **2008**, *130*, 10486–10487.
- [11] P. H. J. Keizers, A. Saragliadis, Y. Hiruma, M. Overhand, M. Ubbink, *J. Am. Chem. Soc.* **2008**, *130*, 14802–14812.
- [12] G. Pintacuda, A. Y. Park, M. A. Keniry, N. E. Dixon, G. Otting, *J. Am. Chem. Soc.* **2006**, *128*, 3696–3702.
- [13] K. D. Brewer, T. Bacaj, A. Cavalli, C. Camilloni, J. D. Swarbrick, J. Liu, A. Zhou, P. Zhou, N. Barlow, J. Xu, A. B. Seven, E. A. Prinslow, R. Voleti, D. Häussinger, A. M. J. J. Bonvin, D. R. Tomchick, M. Vendruscolo, B. Graham, T. C. Südhof, J. Rizo, *Nat. Struct. Mol. Biol.* **2015**, *22*, 555.
- [14] V. Gaponenko, A. S. Altieri, J. Li, R. A. Byrd, *J. Biomol. NMR* **2002**, *24*, 143–148.
- [15] T. Ikegami, L. Verdier, P. Sakhaii, S. Grimme, B. Pescatore, K. Saxena, K. M. Fiebig, C. Griesinger, *J. Biomol. NMR* **2004**, *29*, 339–349.
- [16] J. Wöhnert, K. J. Franz, M. Nitz, B. Imperiali, H. Schwalbe, *J. Am. Chem. Soc.* **2003**, *125*, 13338–13339.
- [17] T. Müntener, J. Kottelat, A. Huber, D. Häussinger, *Bioconjugate Chem.* **2018**, *29*, 3344–3351.
- [18] F. Peters, M. Maestre-Martinez, A. Leonov, L. Kovacic, S. Becker, R. Boelens, C. Griesinger, *J. Biomol. NMR* **2011**, *51*, 329–337.
- [19] D. Joss, R. M. Walliser, K. Zimmermann, D. Häussinger, *J. Biomol. NMR* **2018**, *72*, 29–38.
- [20] T. Müntener, F. Thommen, D. Joss, J. Kottelat, A. Prescimone, D. Häussinger, *Chem. Commun.* **2019**, *55*, 4715–4718.
- [21] R. S. Ranganathan, N. Raju, H. Fan, X. Zhang, M. F. Tweedle, J. F. Desreux, V. Jacques, *Inorg. Chem.* **2002**, *41*, 6856–6866.
- [22] R. S. Ranganathan, R. K. Pillai, N. Raju, H. Fan, H. Nguyen, M. F. Tweedle, J. F. Desreux, V. Jacques, *Inorg. Chem.* **2002**, *41*, 6846–6855.
- [23] A. C. L. Opina, M. Strickland, Y.-S. Lee, N. Tjandra, R. A. Byrd, R. E. Swenson, O. Vasalatiy, *Dalton Trans.* **2016**, *45*, 4673–4687.
- [24] D. Parker, R. S. Dickins, H. Puschmann, C. Crossland, J. A. K. Howard, *Chem. Rev.* **2002**, *102*, 1977–2010.
- [25] L. Dai, C. M. Jones, W. T. K. Chan, T. A. Pham, X. Ling, E. M. Gale, N. J. Rotile, W. C.-S. Tai, C. J. Anderson, P. Caravan, G.-L. Law, *Nat. Commun.* **2018**, *9*, 857.
- [26] S. Tsuboyama, K. Tsuboyama, I. Higashi, M. Yanagita, *Tetrahedron Lett.* **1970**, *11*, 1367–1370.
- [27] M. Frigerio, M. Santagostino, *Tetrahedron Lett.* **1994**, *35*, 8019–8022.
- [28] M. Frigerio, M. Santagostino, S. Sputore, G. Palmisano, *J. Org. Chem.* **1995**, *60*, 7272–7276.
- [29] B. Ravinder, *Tetrahedron Lett.* **2013**, *54*, 4908–4913.
- [30] J.-L. Chen, Y. Zhao, Y.-J. Gong, B.-B. Pan, X. Wang, X.-C. Su, *J. Biomol. NMR* **2018**, *70*, 77–92.
- [31] B. Graham, C. T. Loh, J. D. Swarbrick, P. Ung, J. Shin, H. Yagi, X. Jia, S. Chhabra, N. Barlow, G. Pintacuda, T. Huber, G. Otting, *Bioconjugate Chem.* **2011**, *22*, 2118–2125.
- [32] V. S. Mironov, Y. G. Galyametdinov, A. Ceulemans, C. Görrler-Walrand, K. Binnemans, *J. Chem. Phys.* **2002**, *116*, 4673–4685.
- [33] E. A. Suturina, D. Häussinger, K. Zimmermann, L. Garbuio, M. Yulikov, G. Jeschke, I. Kuprov, *Chem. Sci.* **2017**, *8*, 2751–2757.
- [34] E. A. Suturina, I. Kuprov, *Phys. Chem. Chem. Phys.* **2016**, *18*, 26412–26422.
- [35] H. J. Hogben, M. Krzystyniak, G. T. P. Charnock, P. J. Hore, I. Kuprov, *J. Magn. Reson.* **2011**, *208*, 179–194.
- [36] G. T. P. Charnock, I. Kuprov, *Phys. Chem. Chem. Phys.* **2014**, *16*, 20184–20189.
- [37] R. Ramage, J. Green, T. W. Muir, O. M. Ogunjobi, S. Love, K. Shaw, *Biochem. J.* **1994**, *299*, 151–158.
- [38] J. Sass, F. Cordier, A. Hoffmann, M. Rogowski, A. Cousin, J. G. Omichinski, H. Löwen, S. Grzesiek, *J. Am. Chem. Soc.* **1999**, *121*, 2047–2055.
- [39] W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides, E. D. Laue, *Proteins Struct. Funct. Bioinf.* **2005**, *59*, 687–696.
- [40] D. L. Di Stefano, A. J. Wand, *Biochemistry* **1987**, *26*, 7272–7281.
- [41] B. S. Avvaru, C. U. Kim, K. H. Sippel, S. M. Gruner, M. Agbandje-McKenna, D. N. Silverman, R. McKenna, *Biochemistry* **2010**, *49*, 249–251.
- [42] C. Schmitz, M. J. Stanton-Cook, X.-C. Su, G. Otting, T. Huber, *J. Biomol. NMR* **2008**, *41*, 179.
- [43] F. Neese, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 73–78.
- [44] A. D. Becke, *Phys. Rev. A* **1988**, *38*, 3098–3100.
- [45] J. P. Perdew, *Phys. Rev. B* **1986**, *33*, 8822–8824.
- [46] Y. Takano, K. N. Houk, *J. Chem. Theory Comput.* **2005**, *1*, 70–77.

Manuscript received: April 11, 2019

Revised manuscript received: May 27, 2019

Accepted manuscript online: June 14, 2019

Version of record online: August 12, 2019

CHEMISTRY

A **European** Journal

Supporting Information

A Sterically Overcrowded, Isopropyl-Substituted, Lanthanide-Chelating Tag for Protein Pseudocontact Shift NMR Spectroscopy: Synthesis of its Macrocyclic Scaffold and Benchmarking on Ubiquitin S57 C and hCA II S166 C

Daniel Joss, Maria-Sophie Bertrams, and Daniel Häussinger*^[a]

chem_201901692_sm_miscellaneous_information.pdf

Supporting Information

A sterically overcrowded, isopropyl-substituted lanthanide chelating tag for protein PCS NMR spectroscopy: Synthesis of its macrocyclic scaffold and benchmarking on ubiquitin S57C and hCA II S166C

Daniel Joss,^a Maria-Sophie Bertrams,^a and Daniel Häussinger^{*a}

^a Department of Chemistry, University of Basel, St. Johanns-Ring 19, CH-4056 Basel, daniel.haeussinger@unibas.ch

Table of Content	
HPLC conditions	1
Synthetic procedures	2
HPLC measurements of Ln-P4M4-DOTA	13
Spectra	14
ESI-MS measurements of tagging reactions	24
Shift list comparison	28
Residual dipolar couplings	34
Correlation plots of experimental and back-calculated PCS and RDC	37
Comparison of untagged ubiquitin S57C, Lu-DOTA-M8-(4R4S)-Ub ^{S57C} and Lu-P4M4-DOTA-Ub ^{S57C}	40
Comparison of donor atom positions for dysprosium- and thulium complexes of Ln-DOTA-M8-(4R4S)-SSPy and Ln-P4M4-DOTA obtained from DFT calculations	41
DFT calculations	42
References	43

General Remarks

Unless otherwise stated, reactions were performed under an argon atmosphere and chemicals were used as received without further purification. NMR experiments were performed on Bruker Avance III NMR spectrometers operating at 400, 500 and 600 MHz. Due to unassignable spectra, polyamine compounds longer than a dimeric peptide analog were stabilized with TFA for NMR measurements. ESI-MS spectra were recorded on a Shimadzu LCMS-2020 liquid chromatograph mass spectrometer. HRMS spectra were measured on a Bruker MaXis 4G HiRes ESI Mass Spectrometer.

HPLC conditions

Analytical HPLC measurements were performed on a Shimadzu LC system (LC-20AT prominence liquid chromatograph, SIL-20A HT prominence auto sampler, CTO-20AC prominence column oven, CBM-20A communications bus module, SPD-20A prominence UV/VIS detector ($\lambda = 254$ and 280 nm, LC-20AD prominence liquid chromatograph) combined with a Shimadzu LCMS-2020 liquid chromatograph mass spectrometer. As column for analytical HPLC measurements, a ReproSil-Pur ODS, $3.3 \mu\text{m}$, 150×3 mm, provided by Maisch GmbH was used. Commercial HPLC grade solvents were used and a binary gradient was applied.

Solvent A: Milli-Q water + 0.1% TFA

Solvent B: 90% acetonitrile + 10% Milli-Q water + 0.085% TFA.

HPLC gradient: 95% A (min 0-2), linear gradient 95% A to 100% B (min 2-6), 100% B (min 6-14), linear gradient 100% B to 95% A (min 14-15), 95% A (min 15-22).

Preparative HPLC purification was performed on a Shimadzu LC system (LC-20AT prominence liquid chromatograph, SIL-20A HT prominence auto sampler, CTO-20AC prominence column oven, CBM-20A communications bus module, SPD-20A prominence UV/VIS detector ($\lambda = 254$ and 280 nm, LC-20AD prominence liquid chromatograph) combined with a Shimadzu LCMS-2020 liquid chromatograph mass spectrometer. As column for preparative HPLC purification, a ReproSil-Pur 120 ODS-3, $5 \mu\text{m}$, 150×20 mm, provided by Maisch GmbH was used. Commercial HPLC grade solvents were used and a binary gradient was applied during purification.

Solvent A: Milli-Q water + 0.1% TFA

Solvent B: 90% acetonitrile + 10% Milli-Q water + 0.085% TFA.

HPLC gradient: 95% A (min 0-2), linear gradient 95% A to 100% B (min 2-15), 100% B (min 15-22), linear gradient 100% B to 95% A (min 22-23), 95% A (min 23-25).

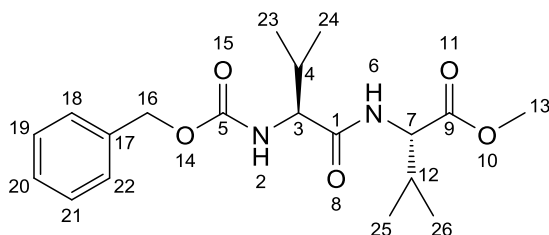
HPLC measurements of protein samples were performed using the direct injection mode on a Shimadzu LC system (LC-20AT prominence liquid chromatograph, SIL-20A HT prominence auto sampler, CTO-20AC prominence column oven, CBM-20A communications bus module, SPD-20A prominence UV/VIS detector ($\lambda = 254$ and 280 nm, LC-20AD prominence liquid chromatograph) combined with a Shimadzu LCMS-2020 liquid chromatograph mass spectrometer. Commercial HPLC grade solvents were used and a binary gradient was applied. MS spectra of proteins were deconvoluted using the Bruker Daltonics DataAnalysis software.

Solvent A: Milli-Q water + 0.1% TFA

Solvent B: 90% acetonitrile + 10% Milli-Q water + 0.085% TFA.

HPLC gradient: isocratic 95% A (min 0-4).

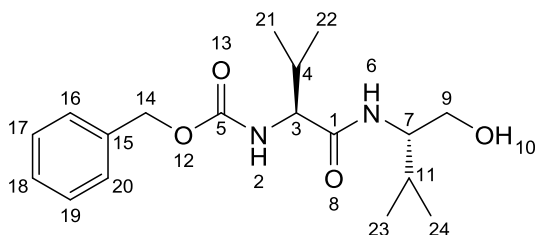
Methyl ((benzyloxy)carbonyl)-L-valyl-L-valinate



DIPEA (39.5 ml, 239 mmol, 3.0 eq.) was slowly added to ((benzyloxy)carbonyl)-L-valine (20.0 g, 79.6 mmol, 1.0 eq.) and methyl L-valinate (13.3 g, 79.6 mmol, 1.0 eq.) in EtOAc (450 ml). Then T3P (61.5 ml, 103 mmol, 1.3 eq., 50% in EtOAc) was added and the mixture was stirred overnight at rt. The organic phase was washed with sat. sodium bicarbonate (150 ml), aq. HCl (150 ml, 0.1 M) and brine (150 ml) and dried over sodium sulfate. The solvent was evaporated to give methyl ((benzyloxy)carbonyl)-L-valyl-L-valinate in quantitative yield.

¹H NMR (400 MHz, DMSO-*d*6) δ 8.10 (d, *J* = 7.8 Hz, 1H, H2), 7.41 – 7.21 (m, 5H, H18-22), 5.02 (s, 2H, H16), 4.15 (dd, *J* = 7.8, 6.3 Hz, 1H, H7), 3.97 (dd, *J* = 9.0, 7.2 Hz, 1H, H3), 3.61 (s, 3H, H13), 2.10 – 1.98 (m, 1H, H12), 1.97 – 1.86 (m, 1H, H4), 0.94 – 0.77 (m, 12H, H23-26).

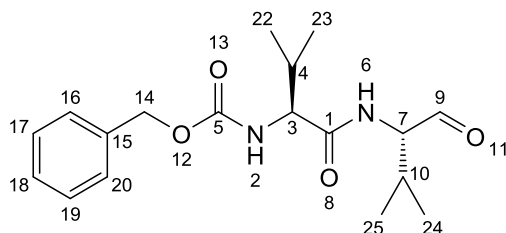
Benzyl ((S)-1-(((S)-1-hydroxy-3-methylbutan-2-yl)amino)-3-methyl-1-oxobutan-2-yl)carbamate



Sodium borohydride (15.6 g, 411 mmol, 5.0 eq.) was added to methyl ((benzyloxy)carbonyl)-L-valyl-L-valinate (30.0 g, 82.3 mmol, 1.0 eq.) in a mixture of THF (450 ml), MeOH (50.0 ml) and water (20.0 ml) and the mixture was stirred overnight at rt. The reaction was quenched by addition of water (200 ml). Organic solvents were evaporated and the aqueous mixture was extracted with DCM (3 x 150 ml). The solvent was evaporated under reduced pressure and the residue subjected to column chromatography (EtOAc, 5.5 x 35 cm) to give benzyl ((S)-1-(((S)-1-hydroxy-3-methylbutan-2-yl)amino)-3-methyl-1-oxobutan-2-yl)carbamate (23.2 g, 83.8%).

¹H NMR (500 MHz, DMSO-*d*6) δ 7.43 (d, *J* = 9.1 Hz, 1H, H2), 7.37 – 7.23 (m, 5H, 16-20), 5.03 (d, *J* = 12.8 Hz, 2H, H14), 4.52 (t, *J* = 5.3 Hz, 1H, H10), 3.85 (dd, *J* = 9.0, 7.3 Hz, 1H, H3), 3.62 – 3.55 (m, 1H, H7), 3.40 – 3.31 (m, 2H + water peak, H9), 2.00 – 1.88 (m, 1H, H4), 1.88 – 1.77 (m, 1H, H11), 0.89 – 0.76 (m, 12H, H21-24).

Benzyl ((S)-3-methyl-1-(((S)-3-methyl-1-oxobutan-2-yl)amino)-1-oxobutan-2-yl)carbamate



Warning: IBX was previously reported to be explosive upon exposure to shock or heat. In our hands, IBX did not explode upon shock exposure and was stable in the temperature ranges employed in this synthetic step and therefore no additional precautions had to be taken. The IBX used in the present work was synthesized in water using Oxone as oxidant.¹

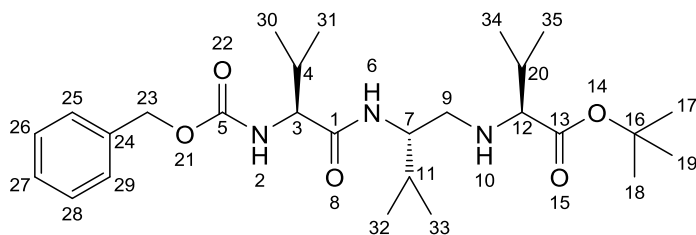
IBX (38.3 g, 137 mmol, 2.0 eq.) was added to ((S)-1-(((S)-1-hydroxy-3-methylbutan-2-yl)amino)-3-methyl-1-oxobutan-2-yl)carbamate (23.0 g, 68.4 mmol, 1.0 eq.), DMSO (4.86 ml, 68.4 mmol, 1.0 eq.) and sodium sulfate (4.86 g, 34.2 mmol, 0.5 eq.) in acetonitrile (250 ml) and the mixture was heated overnight at 80 °C. The mixture was filtrated and solvent was evaporated under reduced pressure. The residue was subjected to column chromatography (EtOAc/cyclohexane, 3:7) to give benzyl ((S)-3-methyl-1-(((S)-3-methyl-1-oxobutan-2-yl)amino)-1-oxobutan-2-yl)carbamate (8.2 g, 35.8%).

¹H NMR (500 MHz, Chloroform-*d*) δ 9.64 (s, 1H, H₉), 9.62 (s, 1H, H₉), 7.40 – 7.29 (m, 5H, H₁₅-20), 6.54 (d, J = 5.3 Hz, 1H, H₂), 6.47 (d, J = 6.4 Hz, 1H, H₂), 5.18 – 5.03 (m, 2H, H₁₄), 4.57 (dd, J = 7.9, 4.4 Hz, 1H, H₇), 4.15 – 4.01 (m, 1H, H₃), 2.39 – 2.25 (m, 1H, H₁₀), 2.23 – 2.08 (m, 1H, H₁₀), 1.23 – 0.79 (m, 12H, H₂₂-H₂₅).

¹³C NMR (126 MHz, Chloroform-*d*) δ 199.5 (C₉), 171.8 (C₁), 171.7 (C₁), 156.8 (C₅), 156.6 (C₅), 136.4 (C₁₅), 128.7 (C_{arom}), 128.4 (C_{arom}), 128.4 (C_{arom}), 128.3 (C_{arom}), 128.2 (C_{arom}), 128.2 (C_{arom}), 67.1 (C₁₄), 63.5 (C₇), 63.4 (C₇), 60.8 (C₃), 31.1 (C₄), 29.2 (C₁₀), 29.1 (C₁₀), 19.4 (C_{22,23}), 19.4 (C_{22,23}), 19.2 (C_{24,25}), 19.2 (C_{24,25}), 17.9 (C_{22,23}), 17.9 (C_{22,23}).

HRMS: [M+H]⁺, C₁₈H₂₇N₂O₄, m/z (calc.) = 335.1965, m/z (meas.) = 335.1971.

tert-butyl ((S)-2-((S)-2-(((benzyloxy)carbonyl)amino)-3-methylbutanamido)-3-methylbutyl)-L-valinate



tert-Butyl *L*-valinate (4.97 g, 71.7 mmol, 1.2 eq.) was added to ((S)-3-methyl-1-(((S)-3-methyl-1-oxobutan-2-yl)amino)-1-oxobutan-2-yl)carbamate (7.99 g, 23.9 mmol, 1.0 eq.) in DCM (125 ml) and the mixture was stirred for 10 min. at rt. Sodium triacetoxyborohydride (15.2 g, 71.7 mmol, 3.0 eq.) was added and the mixture was stirred overnight at rt. The reaction was quenched with water (150 ml). Phases were separated and the aqueous layer was extracted with DCM (3 x 100 ml). The combined organic layers were washed with sat. aq. sodium bicarbonate (150 ml) and brine (150 ml), dried over sodium sulfate and the solvent was evaporated under reduced pressure to give *tert*-butyl ((S)-2-((S)-2-(((benzyloxy)carbonyl)amino)-3-methylbutanamido)-3-methylbutyl)-*L*-valinate (7.2 g, 61.3%).

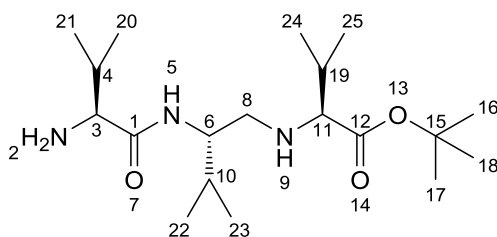
¹H NMR (500 MHz, Acetonitrile-*d*₃) δ 7.59 (d, J = 7.9 Hz, 1H, H₆), 7.41 – 7.29 (m, 5H), 6.55 (d, J = 9.3 Hz, 1H, H₂), 6.21 (d, J = 7.1 Hz, 1H, H₂), 6.10 (d, J = 7.3 Hz, 1H, H₂), 5.97 (d, J = 8.6 Hz, 1H, H₂), 5.90 (d, J = 9.0 Hz, 1H, H₂), 5.15 – 4.99 (m, 2H,

H23), 4.10 (dd, $J = 11.3, 4.0$ Hz, 1H, H7), 4.03 – 3.86 (m, 2H, H3,7), 3.71 (d, $J = 3.9$ Hz, 1H, H12), 3.66 (d, $J = 3.9$ Hz, 1H, H12), 3.26 (dd, $J = 12.9, 2.7$ Hz, 1H, H9), 3.19 (dd, $J = 12.9, 2.8$ Hz, 1H, H9), 3.15 – 3.01 (m, 1H, H9), 2.35 – 2.24 (m, 1H, H20), 2.13 – 1.97 (m, 1H, H4), 1.94 – 1.84 (m, 1H + ACN, H11), 1.51 (s, 9H, H17-19), 1.50 (s, 9H, H17-19), 1.12 – 0.85 (m, 18H, H30-35).

^{13}C NMR (126 MHz, Acetonitrile- d_3) δ 175.5 (C1), 175.1 (C1), 167.6 (C13), 167.5 (C13), 158.0 (C5), 157.7 (C5), 157.3 (C5), 138.2 (C4), 138.0 (C4), 129.5, 129.4, 129.4, 129.4, 128.9 (C27), 128.9 (C27), 128.9 (C27), 128.8 (C27), 128.6 (C_{arom}), 128.6 (C_{arom}), 128.6 (C_{arom}), 128.5 (C_{arom}), 85.8 (C16), 85.7 (C16), 67.4 (C23), 67.3 (C23), 67.2 (C12), 67.0 (C23), 66.6 (C12), 65.1, 62.4, 61.7, 61.6 (C7), 54.7, 54.2, 53.3 (C3), 51.4 (C9), 51.2 (C9), 36.3, 31.6 (C4), 31.0 (C11), 30.9, 30.8 (C11), 30.5, 30.3 (C20), 29.7 (C20), 28.2 (C17-19), 28.2 (C17-19), 28.1 (C17-19), 21.0 – 17.1 (C30-35).

HRMS: $[\text{M}+\text{H}]^+$, $\text{C}_{27}\text{H}_{46}\text{N}_3\text{O}_5$, m/z (calc.) = 492.3432, m/z (meas.) = 492.3431.

***tert*-Butyl ((*S*)-2-((*S*)-2-amino-3-methylbutanamido)-3-methylbutyl)-*L*-valinate**



tert-Butyl ((*S*)-2-((*S*)-2-((benzyloxy)carbonyl)amino)-3-methylbutanamido)-3-methylbutyl)-*L*-valinate (7.18 g, 14.6 mmol, 1.0 eq.) and Palladium 10% on activated charcoal (0.466 g, 4.38 mmol, 0.3 eq.) in MeOH (250 ml) were vigorously stirred overnight at rt under 1 bar H_2 (balloon). The mixture was filtered over Celite and the solvent was evaporated under reduced pressure to give *tert*-butyl ((*S*)-2-((*S*)-2-amino-3-methylbutanamido)-3-methylbutyl)-*L*-valinate (4.6 g, 88.1%).

Species 1:

^1H NMR (500 MHz, Acetonitrile- d_3) δ 8.18 (d, $J = 8.7$ Hz, 1H, H5), 4.12 (dddd, $J = 8.7, 6.2, 3.5, 3.5$ Hz, 1H, H6), 3.86 (d, $J = 3.9$ Hz, 1H, H11), 3.80 (d, $J = 5.3$ Hz, 1H, H3), 3.24 – 3.13 (m, 1H, H8), 2.39 – 2.26 (m, 1H, H19), 2.24 – 2.16 (m, 1H, H4), 1.96 – 1.87 (m, 1H, H10), 1.50 (s, 9H, H16-18), 1.12 – 1.08 (m, 3H, H20-25), 1.04 – 0.97 (m, 9H, H20-25), 0.95 – 0.89 (m, 6H, H20-25).

^{13}C NMR (126 MHz, Acetonitrile- d_3) δ 170.87 (C1), 167.63 (C12), 85.58 (C15), 67.00 (C11), 59.79 (C3), 53.01 (C6), 50.53 (C8), 31.02 (C10), 30.87 (C4), 30.01 (C19), 28.18 or 28.16 (C16-18), 20.12 (C20-25), 20.08 (C20-25), 19.61 (C20-25), 19.08 (C20-25), 19.02 (C20-25), 18.92 (C20-25), 18.68 (C20-25), 18.44 (C20-25), 17.99 (C20-25), 17.97 (C20-25), 17.30 (C20-25), 16.93 (C20-25).

Species 2:

^1H NMR (500 MHz, Acetonitrile- d_3) δ 8.90 (d, $J = 7.0$ Hz, 1H), 3.89 – 3.82 (m, 1H, H6), 3.71 (d, $J = 3.7$ Hz, 1H, H11), 3.70 (d, $J = 6.3$ Hz, 1H, H3), 3.36 (dd, $J = 12.9, 10.6$ Hz, 1H, H8), 3.07 (dd, $J = 12.9, 3.1$ Hz, 1H, H8), 2.39 – 2.26 (m, 1H, H19), 2.29 – 2.20 (m, 1H, H4), 2.12 – 2.01 (m, 1H, H10), 1.51 (s, 9H, H16-18), 1.12 – 1.08 (m, 3H, H20-25), 1.04 – 0.97 (m, 9H, H20-25), 0.95 – 0.89 (m, 6H, H20-25).

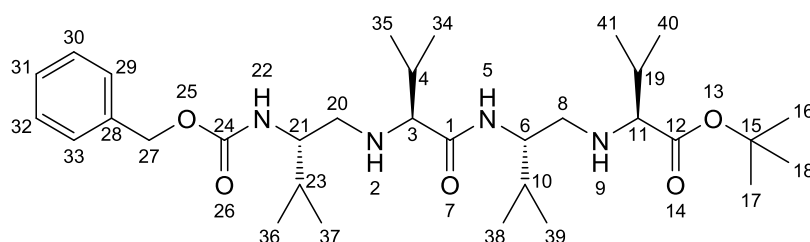
^{13}C NMR (126 MHz, Acetonitrile- d_3) δ 170.35 (C1), 167.60 (C12), 85.67 (C15), 67.67 (C11), 61.15 (C3), 55.59 (C6), 49.79 (C8), 30.87 (C4), 30.60 (C10), 29.93 (C19), 28.18 or 28.16 (C16-18), 20.12 (C20-25), 20.08 (C20-25), 19.61 (C20-25), 19.08 (C20-

25), 19.02 (C20-25), 18.92 (C20-25), 18.68 (C20-25), 18.44 (C20-25), 17.99 (C20-25), 17.97 (C20-25), 17.30 (C20-25), 16.93 (C20-25).

EXSY peaks for amide proton detected in NOESY experiments.

HRMS: $[M+H]^+$, $C_{19}H_{40}N_3O_3$, m/z (calc.) = 358.3064, m/z (meas.) = 358.3071.

***tert*-Butyl ((S)-2-((S)-2-(((S)-2-((benzyloxy)carbonyl)amino)-3-methylbutyl)amino)-3-methylbutanamido)-3-methylbutyl)-L-valinate**



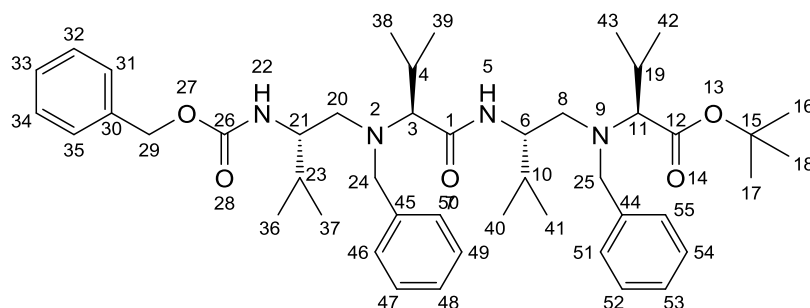
Benzyl (S)-(3-methyl-1-oxobutan-2-yl)carbamate (3.33 g, 30.3 mmol, 1.4 eq.) was added to *tert*-butyl ((S)-2-((S)-2-amino-3-methylbutanamido)-3-methylbutyl)-L-valinate (3.6 g, 10.1 mmol, 1.0 eq.) in DCM (60.0 ml) and the mixture was stirred for 10 min. at rt. Then sodium triacetoxyborohydride (6.42 g, 30.3 mmol, 3.0 eq.) was added and the mixture was stirred overnight at rt. The reaction was quenched with water (100 ml). Phases were separated and the aqueous layer was extracted with DCM (3 x 50 ml). The combined organic layers were washed with sat. aq. sodium bicarbonate (100 ml) and brine (100 ml), dried over sodium sulfate and the solvent was evaporated under reduced pressure to give *tert*-butyl ((S)-2-((S)-2-(((S)-2-((benzyloxy)carbonyl)amino)-3-methylbutyl)amino)-3-methylbutanamido)-3-methylbutyl)-L-valinate in quantitative yield.

¹H NMR (500 MHz, Acetonitrile-*d*3) δ 9.08 (d, J = 9.0 Hz, 1H, H5), 7.43 – 7.26 (m, 5H, H28-29), 6.39 (d, J = 7.6 Hz, 1H, H22), 6.37 (d, J = 7.6 Hz, 1H, H22), 5.20 – 4.99 (m, 2H, H27), 4.19 – 4.10 (m, 1H, H21), 4.00 – 3.93 (m, 1H, H6), 3.78 – 3.65 (m, 3H, H3,11,21), 3.36 – 3.13 (m, 3H, H8,20), 3.12 – 3.01 (m, 1H, H8), 2.39 – 2.27 (m, 1H, H4), 2.13 – 2.03 (m, 1H, H10), 1.92 – 1.75 (m, 2H, H19,23), 1.51 (s, 9H), 1.50 (s, 9H), 1.17 – 0.76 (m, 24H).

¹³C NMR (126 MHz, Acetonitrile-*d*3) δ 169.0 (C1), 167.8 (C12), 167.5 (C12), 159.0 (C24), 158.8 (C24), 137.9 (C28), 137.9 (C28), 129.5 (C_{arom}), 129.4 (C_{arom}), 129.4 (C_{arom}), 129.4 (C_{arom}), 129.0 (C_{arom}), 128.9 (C_{arom}), 128.6 (C_{arom}), 128.6 (C_{arom}), 128.5 (C_{arom}), 85.6 (C15), 85.4 (C15), 67.8 (C27), 67.8 (C21), 67.7 (C11), 67.6 (C11), 67.6 (C27), 67.6 (C27), 55.4 (C6), 54.05, 53.84, 51.41, 51.34, 49.5 (C20), 49.0 (C8), 32.1 (C19), 32.0 (C23), 31.5, 30.9, 30.7 (C10), 30.2 (C4), 30.0, 29.9, 28.2, 28.2, 20.3 (C34-41), 20.1 (C34-41), 19.3 (C34-41), 18.9 (C34-41), 18.9 (C34-41), 18.7 (C34-41), 18.6 (C34-41), 18.4 (C34-41), 18.2 (C34-41), 18.2 (C34-41), 18.1 (C34-41), 18.0 (C34-41), 17.1 (C34-41), 17.1 (C34-41).

HRMS: $[M+H]^+$, $C_{32}H_{57}N_4O_5$, m/z (calc.) = 577.4323, m/z (meas.) = 577.4332.

***tert*-Butyl ((5*S*,8*S*,11*S*,14*S*)-7,13-dibenzyl-5,8,11,14-tetraisopropyl-3,9-dioxo-1-phenyl-2-oxa-4,7,10,13-tetraazapentadecan-15-oate**

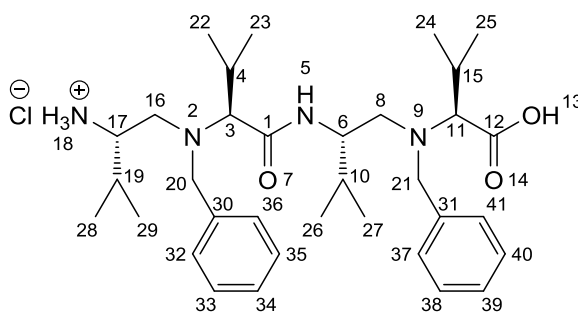


Potassium iodide (0.362 g, 2.18 mmol, 0.2 eq.) was added to benzyl bromide (13.3 ml, 109 mmol, 10.0 eq.), potassium carbonate (15.1 g, 109 mmol, 10.0 eq.) and *tert*-butyl ((*S*)-2-((*S*)-2-(((*S*)-2-((benzyloxy)carbonyl)amino)-3-methylbutyl)amino)-3-methylbutanamido)-3-methylbutyl)-*L*-valinate (6.28 g, 10.9 mmol, 1.0 eq.) in acetonitrile (80 ml) and the mixture was stirred for 3 d at 40 °C. The reaction was quenched by slow addition (over 20 min.) of triethylamine (30.6 ml, 218 mmol, 20.0 eq.). Then the mixture was filtrated and the solvent was evaporated under reduced pressure. The residue was subjected to column chromatography (EtOAc/cyclohexane, 3:7, 1% NEt₃, 5.5 x 35 cm) to give *tert*-butyl ((5*S*,8*S*,11*S*,14*S*)-7,13-dibenzyl-5,8,11,14-tetraisopropyl-3,9-dioxo-1-phenyl-2-oxa-4,7,10,13-tetraazapentadecan-15-oate (5.81 g, 70.4%).

¹H and ¹³C NMR spectra were significantly broadened in deuterated chloroform, acetonitrile, dimethylsulfoxide, acetone, dichloromethane and methanol.

HRMS: [M+H]⁺, C₄₆H₆₉N₄O₅, m/z (calc.) = 757.5262, m/z (meas.) = 757.5250.

***N*-((*S*)-2-((*S*)-2-(((*S*)-2-Amino-3-methylbutyl)(benzyl)amino)-3-methylbutanamido)-3-methylbutyl)-*N*-benzyl-*L*-valine**



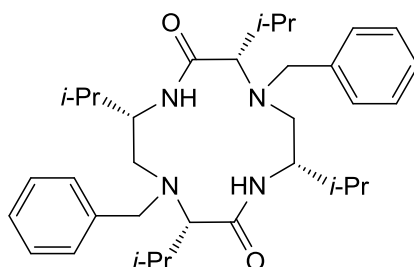
tert-Butyl ((5*S*,8*S*,11*S*,14*S*)-7,13-dibenzyl-5,8,11,14-tetraisopropyl-3,9-dioxo-1-phenyl-2-oxa-4,7,10,13-tetraazapentadecan-15-oate (1.77 g, 2.34 mmol, 1.0 eq.) was dissolved in a solution of HBr in acetic acid (30 ml, 33%) and DCM (20 ml) and the mixture was stirred for 30 min. at rt. The solvent was evaporated under reduced pressure. Aq. HCl (50 ml, 4M) was added and the solvent was evaporated under reduced pressure. This addition/evaporation sequence was repeated with dioxane (1 x 50 ml) and acetonitrile (3 x 50 ml). The crude product *N*-((*S*)-2-((*S*)-2-(((*S*)-2-amino-3-methylbutyl)(benzyl)amino)-3-methylbutanamido)-3-methylbutyl)-*N*-benzyl-*L*-valine was obtained as a mixture of HBr and HCl salts and dried overnight at high vacuum at 40 °C. The product was checked for absence of acetic acid by NMR and then used as such for cyclization.

¹H NMR (500 MHz, Acetonitrile-*d*₃) δ 10.71 (s, 1H, H13), 7.70 – 7.17 (m, 13H, H32-41,18), 3.98 (d, J = 13.5 Hz, 1H, H21), 3.90 (d, J = 14.0 Hz, 1H, H20), 3.80 – 3.65 (m, 3H, H20,21,3), 3.29 – 3.08 (m, 3H, H6,8,11,16), 2.99 – 2.81 (m, 3H, H8,16,17), 2.67 – 2.51 (m, 1H, H4), 2.22 – 2.11 (m, 1H, H15), 1.88 – 1.72 (m, 2H, H10,19), 1.41 – 0.53 (m, 24H, H22-29).

¹³C NMR (126 MHz, Acetonitrile-*d*3) δ 178.7 (C12), 171.2 (C1), 170.5 (C1), 140.2 (C31), 139.9 (C30), 132.1 (C_{arom}), 132.0 (C_{arom}), 130.9 (C_{arom}), 130.7 (C_{arom}), 130.4 (C_{arom}), 130.3 (C_{arom}), 130.2 (C_{arom}), 130.0 (C_{arom}), 129.9 (C_{arom}), 129.9 (C_{arom}), 129.8 (C_{arom}), 129.8 (C_{arom}), 129.6 (C_{arom}), 129.5 (C_{arom}), 128.4 (C_{arom}), 128.3 (C_{arom}), 73.8 (C3), 71.2 (C11), 69.9 (C17), 69.6, 59.1 (C20), 57.9, 57.6, 57.2, 56.5 (C21), 54.9, 54.4 (C6), 53.9, 52.3 (C8), 51.4 (C16), 31.0 (C10), 30.9, 30.1, 29.8, 29.3, 29.1 (C19), 29.0, 29.0 (C15), 28.0, 27.2 (C4), 25.6, 23.5, 21.9 (C_{isoprop}), 21.1 (C_{isoprop}), 21.0 (C_{isoprop}), 20.6 (C_{isoprop}), 20.4 (C_{isoprop}), 19.9 (C_{isoprop}), 19.6 (C_{isoprop}), 19.1 (C_{isoprop}), 18.7 (C_{isoprop}), 18.7 (C_{isoprop}), 18.1 (C_{isoprop}), 18.1 (C_{isoprop}), 17.9 (C_{isoprop}), 17.7.

HRMS: [M+H]⁺, C₃₄H₅₅N₄O₃, m/z (calc.) = 567.4269, m/z (meas.) = 567.4280.

(3S,6S,9S,12S)-4,10-Dibenzyl-3,6,9,12-tetraisopropyl-1,4,7,10-tetraazacyclododecane-2,8-dione

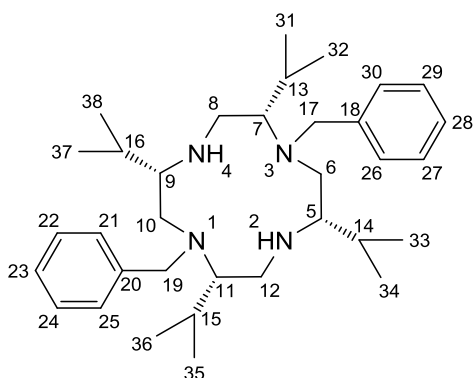


HATU (1.06 g, 2.80 mmol, 2.0 eq.) was added in one portion to a vigorously stirred solution of *N*-((*S*)-2-(((*S*)-2-amino-3-methylbutyl)(benzyl)amino)-3-methylbutanamido)-3-methylbutyl)-*N*-benzyl-*L*-valine HBr/HCl salt (0.950 g, 1.40 mmol, 1.0 eq.) and DIPEA (2.31 ml, 14.0 mmol, 10.0 eq.) in acetonitrile (950 ml) and the mixture was stirred for 5 h at rt. The solvent was evaporated under reduced pressure and the residue was subjected to column chromatography (EtOAc, 1% NEt₃) to give (3*S*,6*S*,9*S*,12*S*)-4,10-dibenzyl-3,6,9,12-tetraisopropyl-1,4,7,10-tetraazacyclododecane-2,8-dione (0.53 g, 65.3% over two steps).

¹H and ¹³C NMR spectra were significantly broadened in deuterated chloroform, acetonitrile and dimethylsulfoxide.

HRMS: [M+H]⁺, C₃₄H₅₃N₄O₂, m/z (calc.) = 549.4163, m/z (meas.) = 549.4174.

(2S,5S,8S,11S)-1,7-Dibenzyl-2,5,8,11-tetraisopropyl-1,4,7,10-tetraazacyclododecane



TMSCl (1.10 ml, 8.64 mmol, 5.5 eq.) was added to (3*S*,6*S*,9*S*,12*S*)-4,10-dibenzyl-3,6,9,12-tetraisopropyl-1,4,7,10-tetraazacyclododecane-2,8-dione (862 mg, 1.57 mmol, 1.0 eq.) in DCM (15.0 ml) and the mixture was stirred for

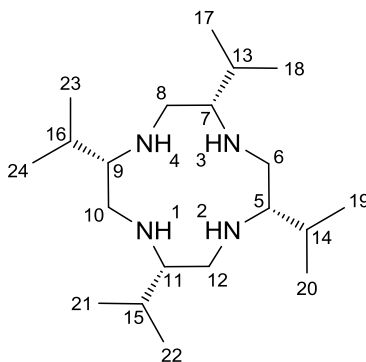
30 min. at 0 °C. Then lithium aluminium hydride (8.64 ml, 17.3 mmol, 11.0 eq.) was added and the mixture was warmed to rt over the course of 30 min. Then the mixture was stirred for 2.5 h at rt. HPLC-MS indicated the presence of starting material in the reaction mixture. Lithium aluminium hydride (4.32 ml, 8.64 mmol, 5.5 eq.) was added and the mixture was stirred for 2 d at rt. The reaction was quenched with aq. KOH (2.83 ml, 4.0 M) and subsequently dried with excess sodium sulfate. The mixture was filtered through a plug of sodium sulfate and the filter cake was washed with DCM/ethanol (9:1, 250 ml). The solvent was evaporated and the residue subjected to column chromatography (DCM/ethanol, 8:2, 2% NEt₃) to give (2*S*,5*S*,8*S*,11*S*)-1,7-dibenzyl-2,5,8,11-tetraisopropyl-1,4,7,10-tetraazacyclododecane (752 mg, 92.0%).

¹H NMR (500 MHz, Acetonitrile-*d*₃) δ 7.45 – 7.32 (m, 10H, H21-30), 3.99 (d, J = 13.6 Hz, 2H, H17), 3.69 (d, J = 13.6 Hz, 2H, H17), 3.56 (ddd, J = 9.5 Hz, 1H, H5,9), 3.15 – 3.09 (m, 1H, H7,11), 2.87 – 2.77 (m, 4H, H6,8,10,12), 2.08 – 2.00 (m, 1H, H13,15), 2.00 – 1.93 (m, 1H, H14,16), 1.12 – 0.83 (m, 24H).

¹³C NMR (126 MHz, Acetonitrile-*d*₃) δ 137.54 (C18,20), 130.60 (C21,25,26,30), 130.07 (C22,24,27,29), 129.14 (C23,28), 62.88 (C7,11), 58.39 (C5,9), 56.11 (C17,19), 45.83 (C6,10), 38.52 (C8,12), 27.55 (C14,16), 25.70 (C13,15), 23.09 (C31,32,35,36), 21.30 (C33,34,37,38), 19.27 (C33,34,37,38), 18.63 (C31,32,35,36).

HRMS: [M+H]⁺, C₃₄H₅₇N₄, m/z (calc.) = 521.4578, m/z (meas.) = 521.4587.

(2*S*,5*S*,8*S*,11*S*)-2,5,8,11-Tetraisopropyl-1,4,7,10-tetraazacyclododecane



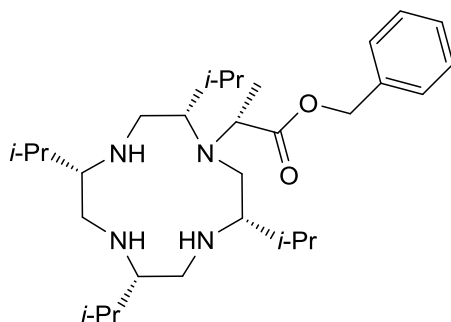
Ammonium formate (1.65 g, 26.2 mmol, 20.0 eq.) was added to (2*S*,5*S*,8*S*,11*S*)-1,7-dibenzyl-2,5,8,11-tetraisopropyl-1,4,7,10-tetraazacyclododecane (682 mg, 1.31 mmol, 1.0 eq.) and palladium hydroxide on carbon (269 mg) in EtOH (20.0 ml) and the mixture was stirred at reflux for 5 h. Then ammonium formate (1.65 g, 26.2 mmol, 20.0 eq.) was added and the mixture was stirred for 2 h at reflux. Ammonium formate (0.83 g, 13.1 mmol, 10.0 eq.) was added and the mixture was stirred overnight at reflux. The mixture was filtered through Celite and the solvent was evaporated under reduced pressure. The crude compound was dissolved in DCM (100 ml) and KOH (100 ml) was added. Phases were separated and the aqueous layer was extracted with DCM (3x 50 ml). The combined organic layers were dried over sodium sulfate and the solvent was evaporated under reduced pressure to give (2*S*,5*S*,8*S*,11*S*)-2,5,8,11-tetraisopropyl-1,4,7,10-tetraazacyclododecane (345 mg, 77.3%).

¹H NMR (500 MHz, Chloroform-*d*) δ 2.62 (dd, J = 12.9, 3.0 Hz, 4H, H6,8,10,12eq.), 2.45 (dd, J = 12.9, 8.8 Hz, 4H, H6,8,10,12ax.), 2.32 (ddd, J = 8.7, 4.9, 3.0 Hz, 4H, H5,7,9,11), 1.86 – 1.75 (m, 4H, H13-16), 0.92 (d, J = 6.9 Hz, 12H, H17-24), 0.85 (d, J = 6.9 Hz, 12H, H17-24).

¹³C NMR (126 MHz, Chloroform-*d*) δ 58.74 (C5,7,9,11), 45.52 (C6,8,10,12), 28.47 (C13,14,15,16), 20.16 (C17,19,21,23), 17.20 (C18,20,22,24).

HRMS: [M+H]⁺, C₂₀H₄₅N₄, m/z (calc.) = 341.3639, m/z (meas.) = 341.3643.

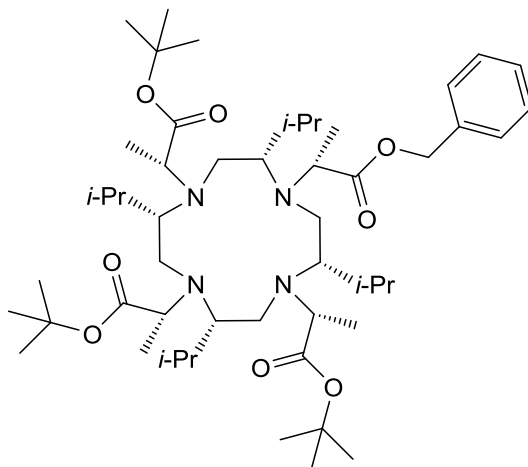
Benzyl (R)-2-((2S,5S,8S,11S)-2,5,8,11-tetraisopropyl-1,4,7,10-tetraazacyclododecan-1-yl)propanoate



Benzyl (S)-2-(((trifluoromethyl)sulfonyl)oxy)propanoate (183 mg, 0.587 mmol, 1.0 eq.) was added dropwise over 30 min. to (2S,5S,8S,11S)-2,5,8,11-tetraisopropyl-1,4,7,10-tetraazacyclododecane (200 mg, 0.587 mmol, 1.0 eq.) in acetonitrile (10 ml) and the mixture was stirred at rt for 2 h. The reaction was quenched by addition of an excess triethylamine, the solvent was evaporated under reduced pressure and the residue subjected to preparative HPLC to give benzyl (R)-2-((2S,5S,8S,11S)-2,5,8,11-tetraisopropyl-1,4,7,10-tetraazacyclododecan-1-yl)propanoate (152 mg, 51.5%).

HRMS: [M+H]⁺, C₃₀H₅₅N₄O₂, m/z (calc.) = 503.4320, m/z (meas.) = 503.4328.

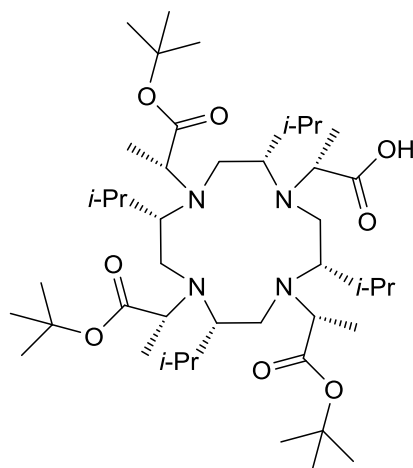
Tri-*tert*-butyl 2,2',2''-((2S,5S,8S,11S)-10-((R)-1-(benzyloxy)-1-oxopropan-2-yl)-2,5,8,11-tetraisopropyl-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)(2R,2'R,2''R)-tripropionate



tert-Butyl (S)-2-(((trifluoromethyl)sulfonyl)oxy)propanoate (1.26 g, 4.53 mmol, 15.0 eq.) in acetonitrile (1.5 ml) was added to benzyl (R)-2-((2S,5S,8S,11S)-2,5,8,11-tetraisopropyl-1,4,7,10-tetraazacyclododecan-1-yl)propanoate (152 mg, 0.302 mmol, 1.0 eq.) and potassium carbonate (626 mg, 4.53 mmol, 15.0 eq.) in acetonitrile (1.5 ml) and the mixture was stirred at overnight. The reaction was quenched by addition of an excess triethylamine, the solvent was evaporated under reduced pressure and the residue subjected to preparative HPLC to give tri-*tert*-butyl 2,2',2''-((2S,5S,8S,11S)-10-((R)-1-(benzyloxy)-1-oxopropan-2-yl)-2,5,8,11-tetraisopropyl-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)(2R,2'R,2''R)-tripropionate (40.0 mg, 14.9%).

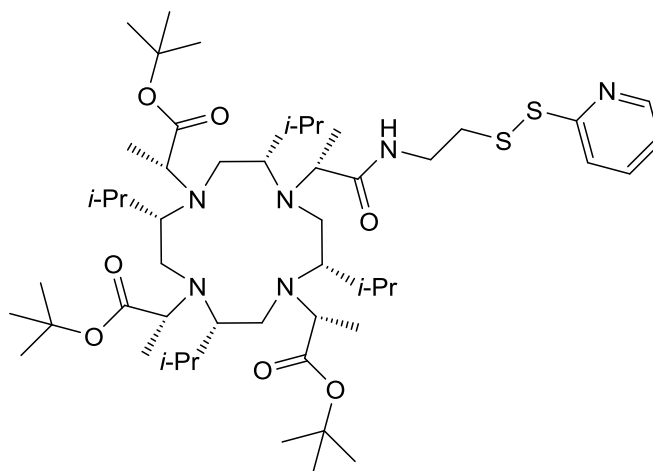
HRMS: [M+H]⁺, C₅₁H₉₁N₄O₈, m/z (calc.) = 887.6831, m/z (meas.) = 887.6818.

(R)-2-((2S,5S,8S,11S)-4,7,10-Tris((R)-1-(tert-butoxy)-1-oxopropan-2-yl)-2,5,8,11-tetraisopropyl-1,4,7,10-tetraazacyclododecan-1-yl)propanoic acid



Palladium 10% on carbon (10.0 mg) was added to tri-*tert*-butyl 2,2',2''-((2S,5S,8S,11S)-10-((R)-1-(benzyloxy)-1-oxopropan-2-yl)-2,5,8,11-tetraisopropyl-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)(2R,2'R,2''R)-tripropionate (39.0 mg, 0.044 mmol, 1.0 eq.) in MeOH (5 ml) and the mixture was stirred vigorously at rt under a pressure of 20 bar hydrogen for 3 h. The mixture was filtered over Celite and the solvent was evaporated under reduced pressure to give (R)-2-((2S,5S,8S,11S)-4,7,10-tris((R)-1-(*tert*-butoxy)-1-oxopropan-2-yl)-2,5,8,11-tetraisopropyl-1,4,7,10-tetraazacyclododecan-1-yl)propanoic acid (32.0 mg, 91.2%), which was used as such.

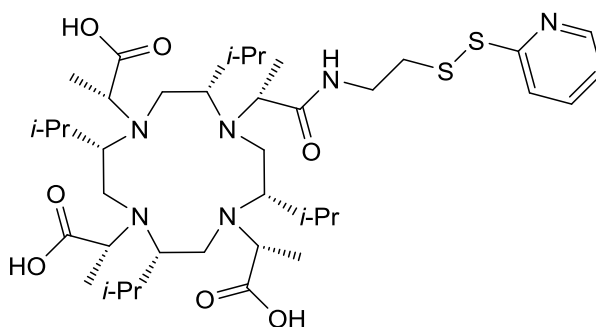
Tri-*tert*-butyl 2,2',2''-((2S,5S,8S,11S)-2,5,8,11-tetraisopropyl-10-((R)-1-oxo-1-((2-(pyridin-2-yl)disulfaneyl)ethyl)amino)propan-2-yl)-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)(2R,2'R,2''R)-tripropionate



HATU (88.7 mg, 0.233 mmol, 6.0 eq.) was added to (R)-2-((2S,5S,8S,11S)-4,7,10-tris((R)-1-(*tert*-butoxy)-1-oxopropan-2-yl)-2,5,8,11-tetraisopropyl-1,4,7,10-tetraazacyclododecan-1-yl)propanoic acid (31.0 mg, 0.0389 mmol, 1.0 eq.), 2-(pyridin-2-yl)disulfaneyl)ethan-1-amine (43.3 mg, 0.194 mmol, 5.0 eq.) and DIPEA (0.064 ml, 0.389 mmol, 10.0 eq.) in DMF (2.0 ml) and the mixture was stirred at rt for 1 h. The mixture was filtered and subjected to preparative HPLC to give tri-*tert*-butyl 2,2',2''-((2S,5S,8S,11S)-2,5,8,11-tetraisopropyl-10-((R)-1-oxo-1-((2-(pyridin-2-yl)disulfaneyl)ethyl)amino)propan-2-yl)-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)(2R,2'R,2''R)-tripropionate (10.0 mg, 26.6%).

HRMS: [M+H]⁺. C₅₁H₉₃N₆O₇S₂, m/z (calc.) = 965.6542, m/z (meas.) = 965.6543.

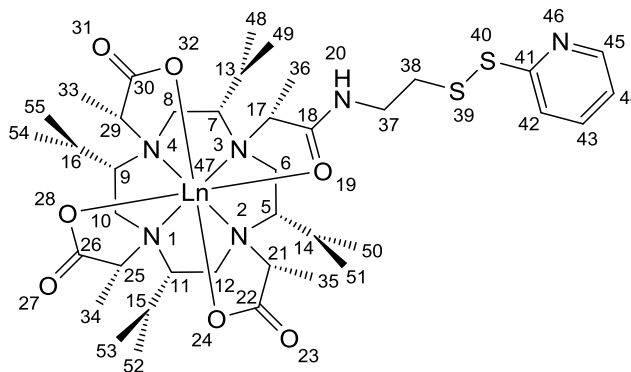
(2*R*,2'*R*,2''*R*)-2,2',2''-((2*S*,5*S*,8*S*,11*S*)-2,5,8,11-Tetraisopropyl-10-((*R*)-1-oxo-1-((2-(pyridin-2-yl)disulfaneyl)ethyl)amino)propan-2-yl)-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)tripropionic acid



Tri-*tert*-butyl 2,2',2''-((2*S*,5*S*,8*S*,11*S*)-2,5,8,11-tetraisopropyl-10-((*R*)-1-oxo-1-((2-(pyridin-2-yl)disulfaneyl)ethyl)amino)propan-2-yl)-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)(2*R*,2'*R*,2''*R*)-tripropionate (9.0 mg, 0.0104 mmol, 1.0 eq.) was dissolved in a mixture of TFA (1.84 ml), thioanisole (0.12 ml) and water (0.04 ml) and stirred at rt for 8 h. The solvent was evaporated under reduced pressure and the residue subjected to preparative HPLC to give (2*R*,2'*R*,2''*R*)-2,2',2''-((2*S*,5*S*,8*S*,11*S*)-2,5,8,11-tetraisopropyl-10-((*R*)-1-oxo-1-((2-(pyridin-2-yl)disulfaneyl)ethyl)amino)propan-2-yl)-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)tripropionic acid (5.0 mg, 60.3%).

HRMS: [M+H]⁺, C₃₉H₆₉N₆O₇S₂, m/z (calc.) = 797.4664, m/z (meas.) = 797.4651.

Ln-P4M4-DOTA



Ln(OTf)₃ (0.95 eq.) was added to (2*R*,2'*R*,2''*R*)-2,2',2''-((2*S*,5*S*,8*S*,11*S*)-2,5,8,11-tetraisopropyl-10-((*R*)-1-oxo-1-((2-(pyridin-2-yl)disulfaneyl)ethyl)amino)propan-2-yl)-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)tripropionic acid (1.5 mg, 0.00188 mmol, 1.0 eq.) in aq. ammonium acetate (1.5 ml, 100 mM) and the mixture was heated to 80 °C overnight. The mixture was subjected to preparative HPLC to give Ln-P4M4-DOTA (43.8-55.5%).

Λ(δδδδ)-Lu-P4M4-DOTA

¹H NMR (600 MHz, Deuterium Oxide) δ 8.43 (s, 1H, H45), 7.95 – 7.86 (m, 2H, H42,43), 7.37 (dd, J = 7.2, 5.2 Hz, 1H, H44), 4.10 (q, J = 7.4 Hz, 1H, H17), 3.86 (ddd, J = 13.4, 7.6, 5.1 Hz, 1H, H37), 3.81 (q, J = 7.6 Hz, 1H, H25), 3.79 (q, J = 7.6 Hz, 1H, H21), 3.67 (q, J = 7.3 Hz, 1H), 3.41 (ddd, J = 14.1, 5.6, 5.6 Hz, 1H), 3.06 – 2.93 (m, 8H), 2.85 (dd, J = 10.7, 3.8 Hz, 1H, H11), 2.83 – 2.75 (m, 2H, H7,9), 2.75 – 2.70 (m, 1H, H5), 2.54 – 2.36 (m, 4H, H13,14,15,16), 1.71 (d, J = 7.3 Hz, 3H, H36), 1.59 – 1.49 (m, 8H, H33,34,35), 1.09 – 1.03 (m, 12H, H49,51,53,55), 1.03 – 0.97 (m, 12H, H48,50,52,54).

¹³C NMR (151 MHz, Deuterium Oxide, extracted from HMQC and HMBC) δ 183.0 (C26), 182.8 (C30), 182.8 (C22), 181.0 (C18), 148.3 (C45), 139.8 (C43), 122.44 (C44), 122.0 (C42), 71.1 (C5), 70.7 (C7), 70.4 (C11), 70.0 (C9), 68.6 (C21), 67.9 (C29), 67.9 (C25), 65.7 (C17), 42.1 (C6), 42.0 (C10), 42.0 (C12), 41.9 (C8), 39.1 (C37), 37.2 (C38), 27.2 (C13), 27.1 (C14), 27.0 (C15), 27.0 (C16), 23.6 (C49,51,53,55), 17.1 (C48,50,52,54), 14.9 (C36), 14.1 (C35), 13.7 (C34), 13.6 (C33).

Ln = Lu, **HRMS**: [M+H]⁺, C₃₉H₆₆LuN₆O₇S₂, m/z (calc.) = 969.3837, m/z (meas.) = 969.3827.

Ln = Tm, **HRMS**: [M+H]⁺, C₃₉H₆₆N₆O₇S₂Tm, m/z (calc.) = 963.3771, m/z (meas.) = 963.3760.

Ln = Dy, **HRMS**: [M+H]⁺, C₃₉H₆₆DyN₆O₇S₂, m/z (calc.) = 958.3721, m/z (meas.) = 958.3705.

HPLC measurements of Ln-P4M4-DOTA

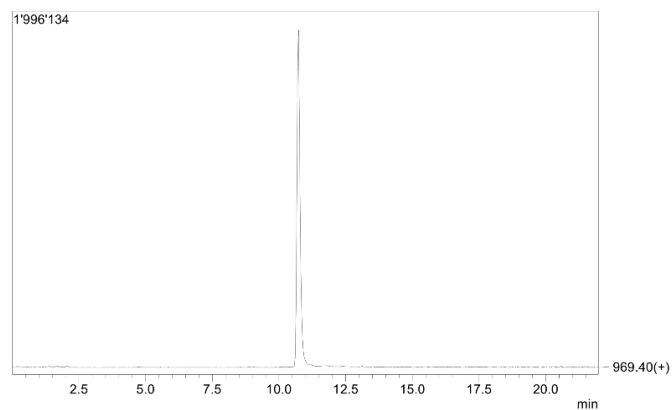


Figure S1: HPLC-MS trace of Lu-P4M4-DOTA.

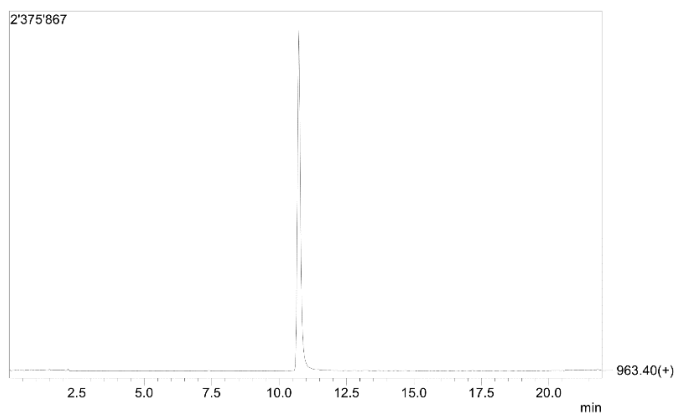


Figure S2: HPLC-MS trace of Tm-P4M4-DOTA.

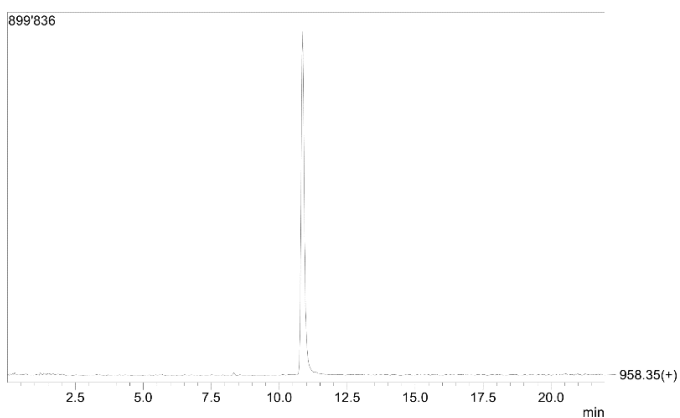
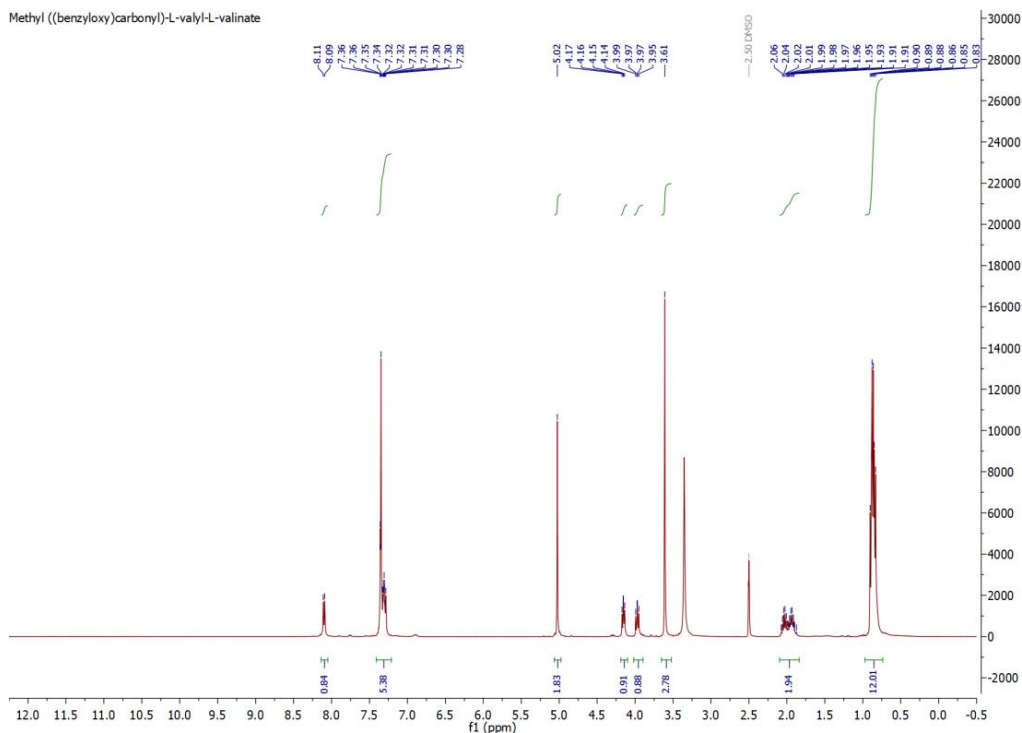


Figure S3: HPLC-MS trace of Dy-P4M4-DOTA.

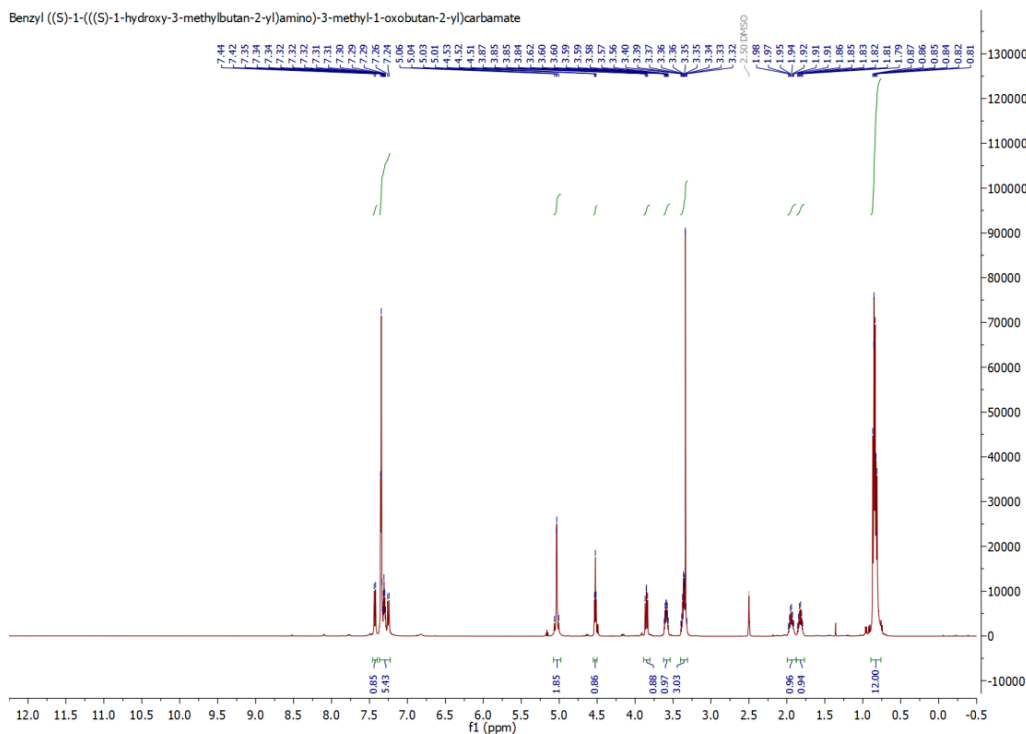
Methyl ((benzyloxy)carbonyl)-L-valyl-L-valinate

¹H NMR



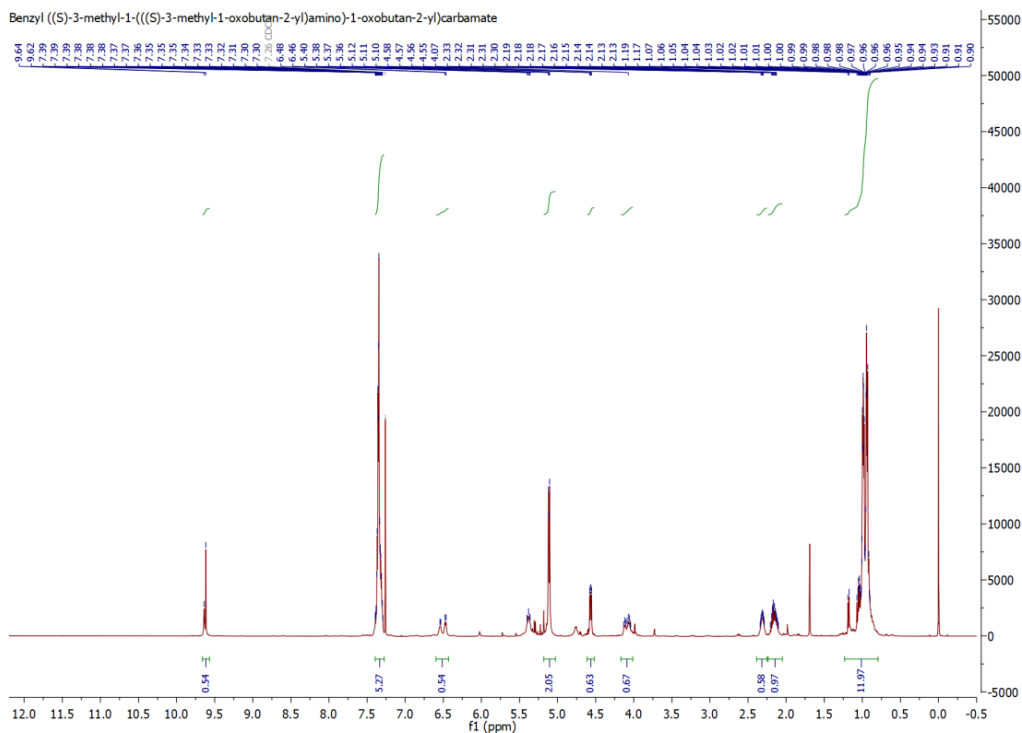
Benzyl ((S)-1-(((S)-1-hydroxy-3-methylbutan-2-yl)amino)-3-methyl-1-oxobutan-2-yl)carbamate

¹H NMR

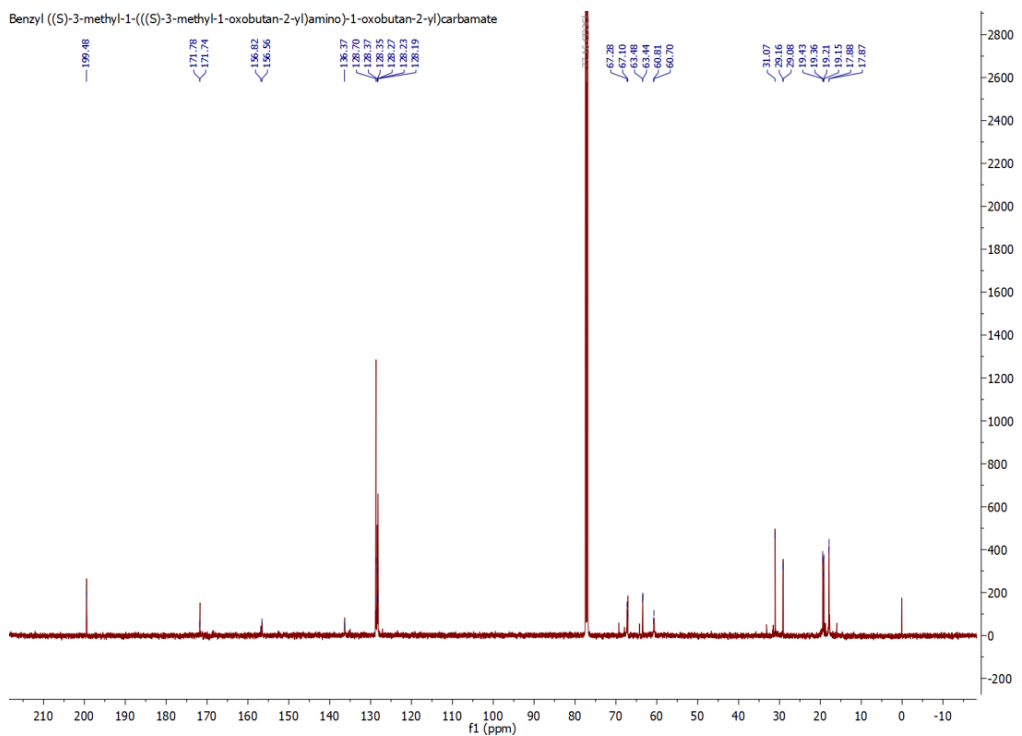


Benzyl ((S)-3-methyl-1-(((S)-3-methyl-1-oxobutan-2-yl)amino)-1-oxobutan-2-yl)carbamate

¹H NMR

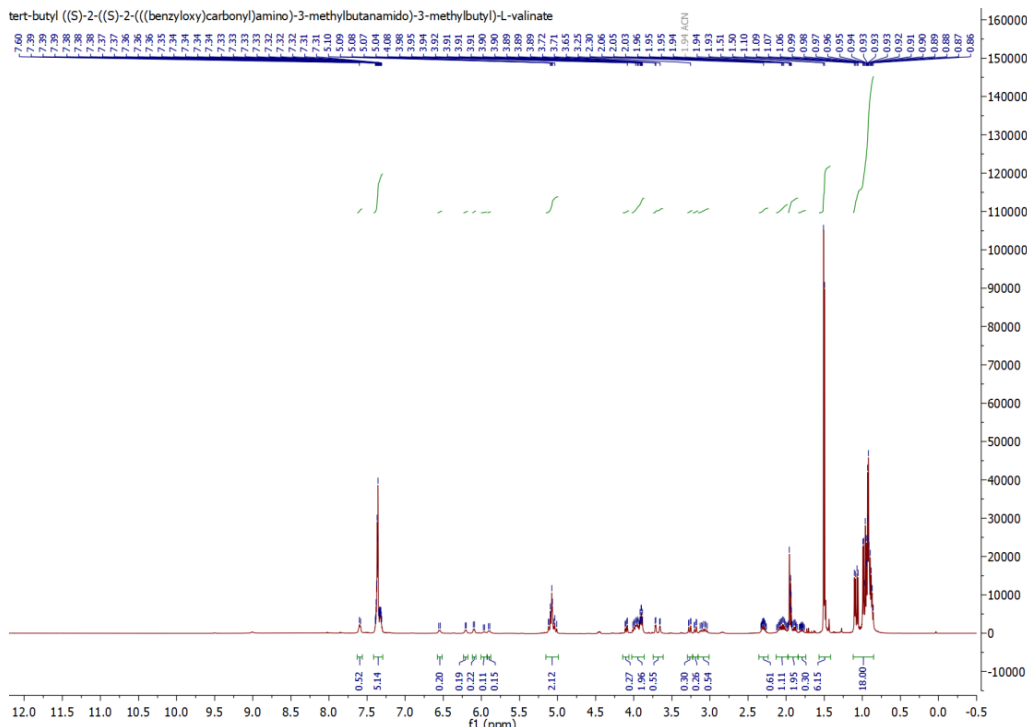


¹³C NMR

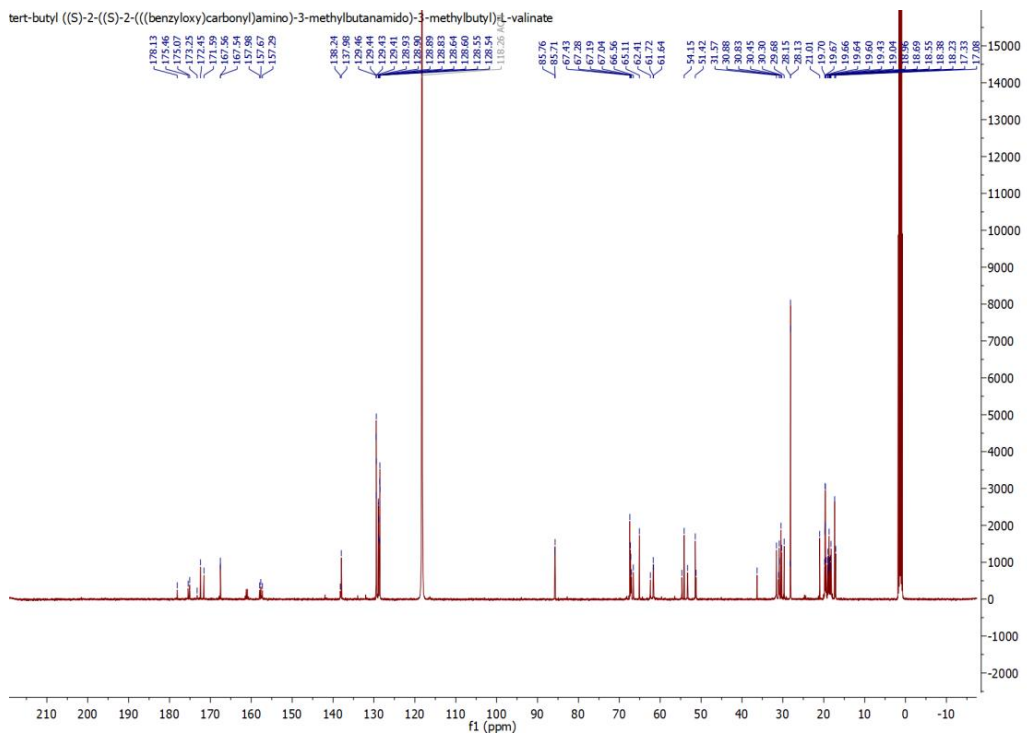


tert-butyl ((S)-2-((S)-2-(((benzyloxy)carbonyl)amino)-3-methylbutanamido)-3-methylbutyl)-L-valinate

¹H NMR

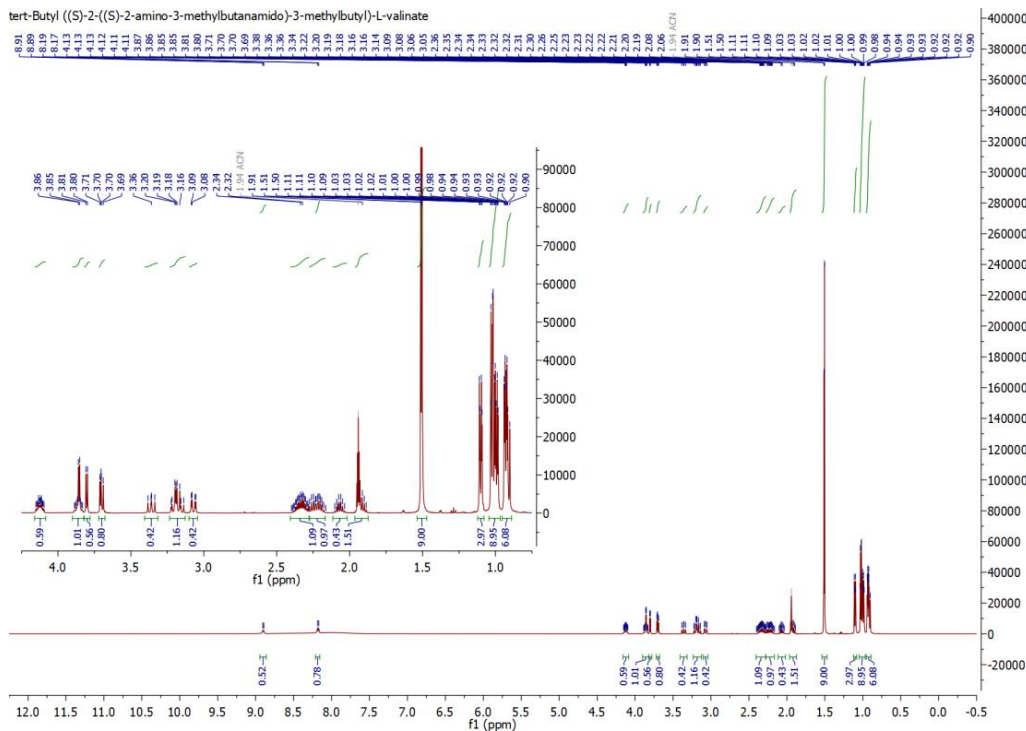


¹³C NMR

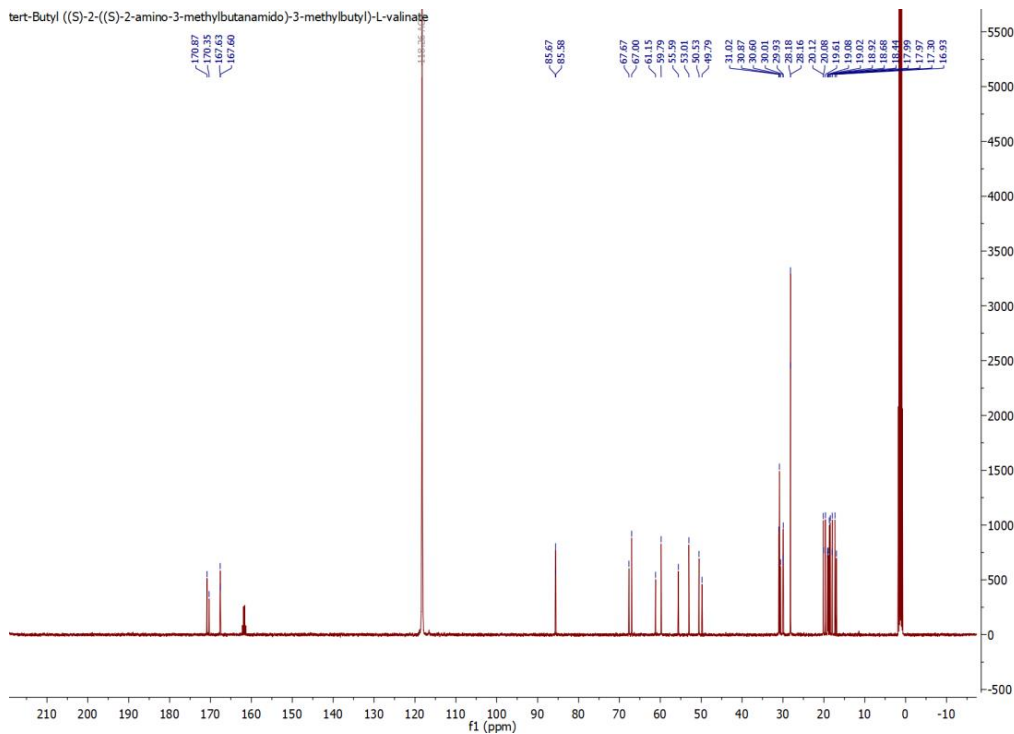


tert-Butyl ((S)-2-((S)-2-amino-3-methylbutanamido)-3-methylbutyl)-L-valinate

¹H NMR

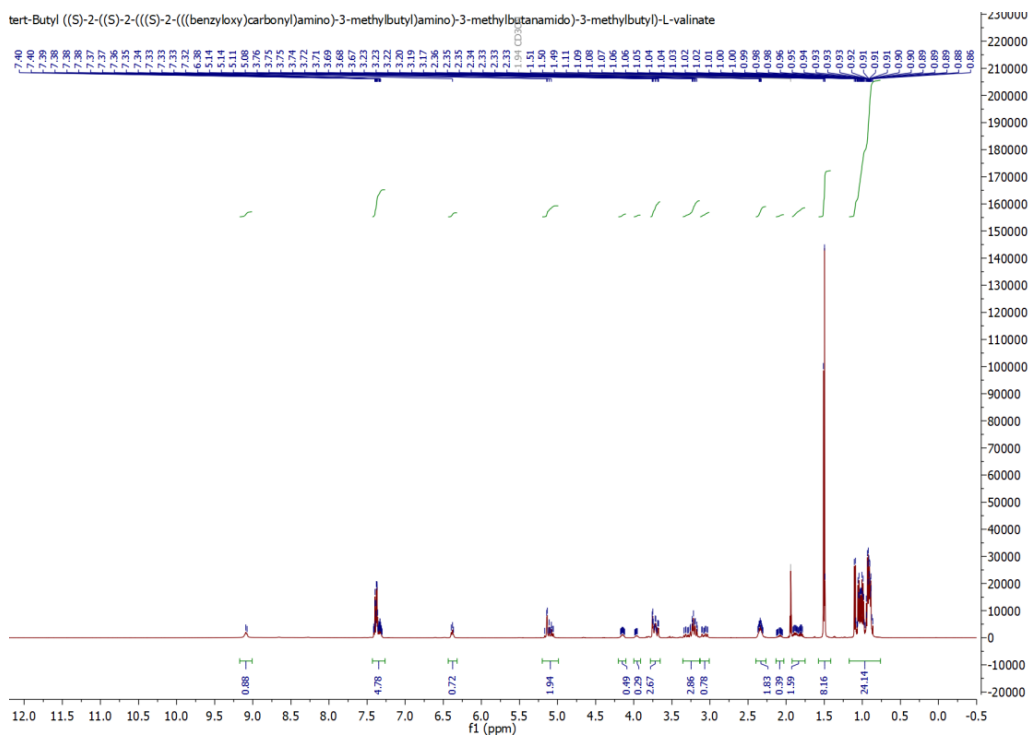


¹³C NMR

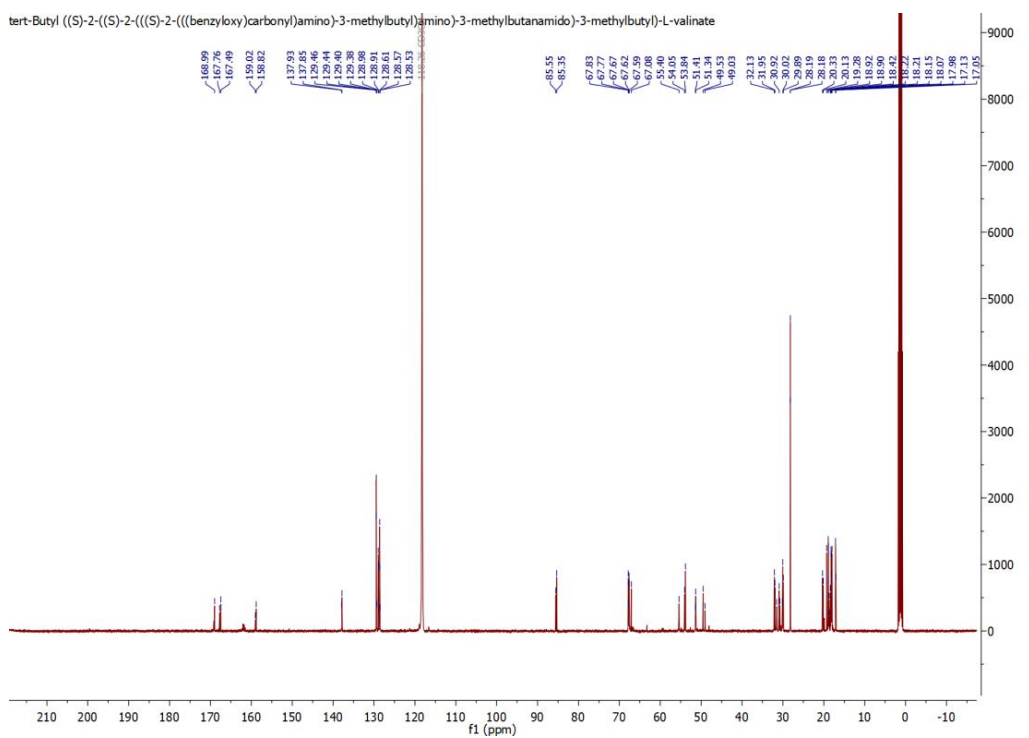


tert-Butyl ((S)-2-((S)-2-(((S)-2-(((benzyloxy)carbonyl)amino)-3-methylbutyl)amino)-3-methylbutanamido)-3-methylbutyl)-L-valinate

¹H NMR

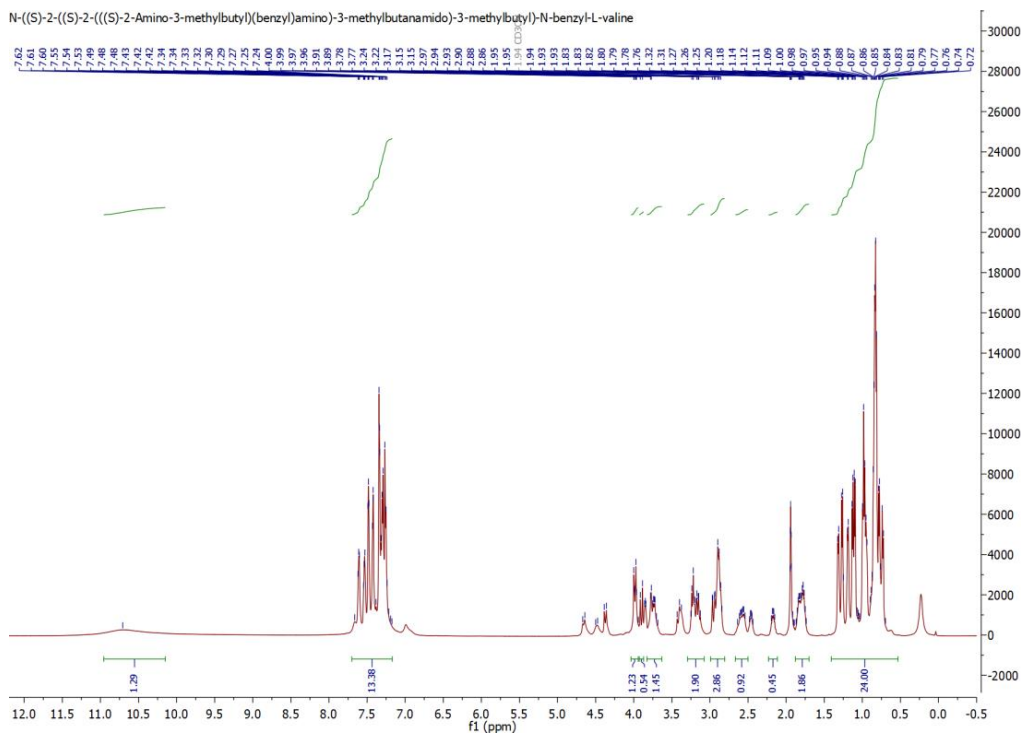


¹³C NMR

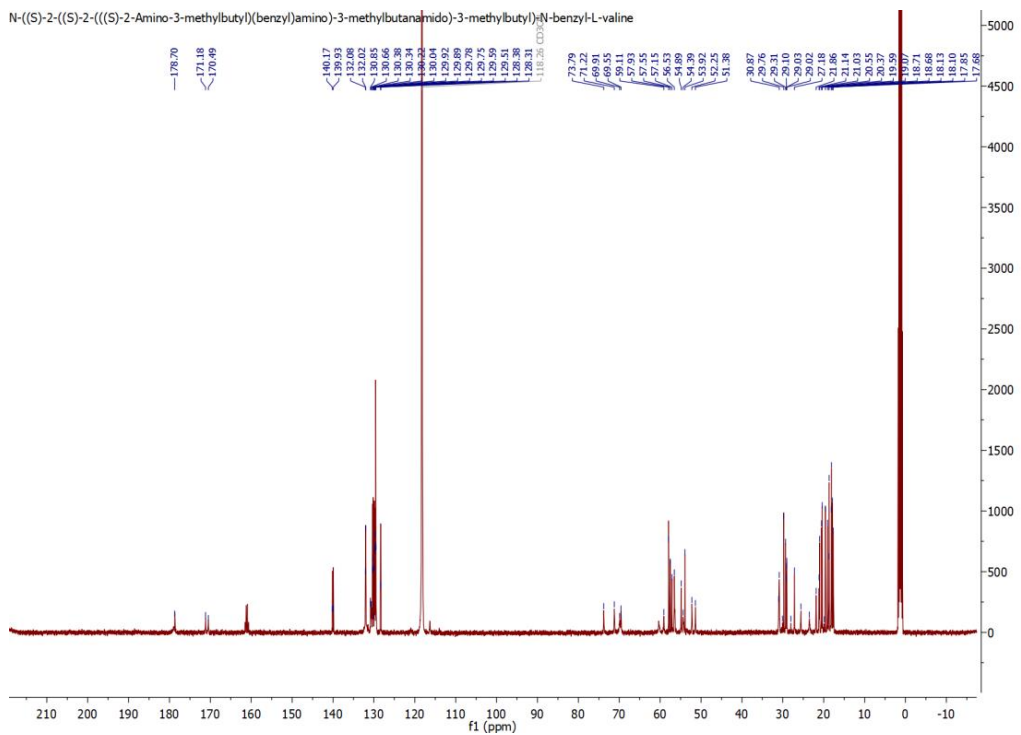


N-((S)-2-((S)-2-(((S)-2-Amino-3-methylbutyl)(benzyl)amino)-3-methylbutanamido)-3-methylbutyl)-N-benzyl-L-valine

¹H NMR

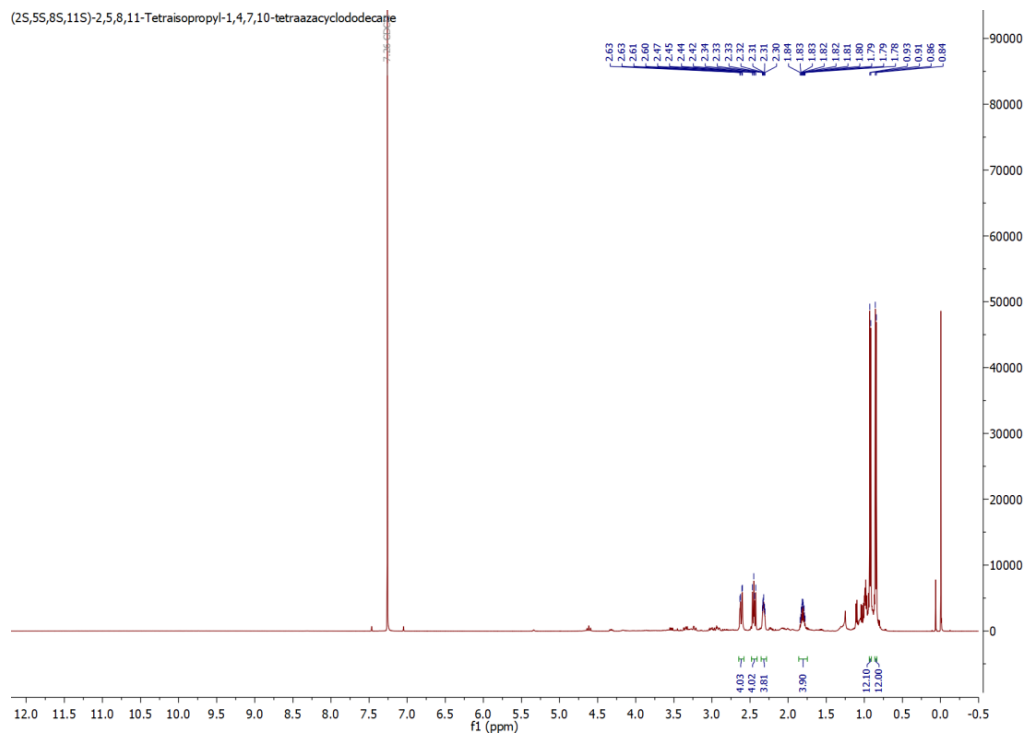


¹³C NMR

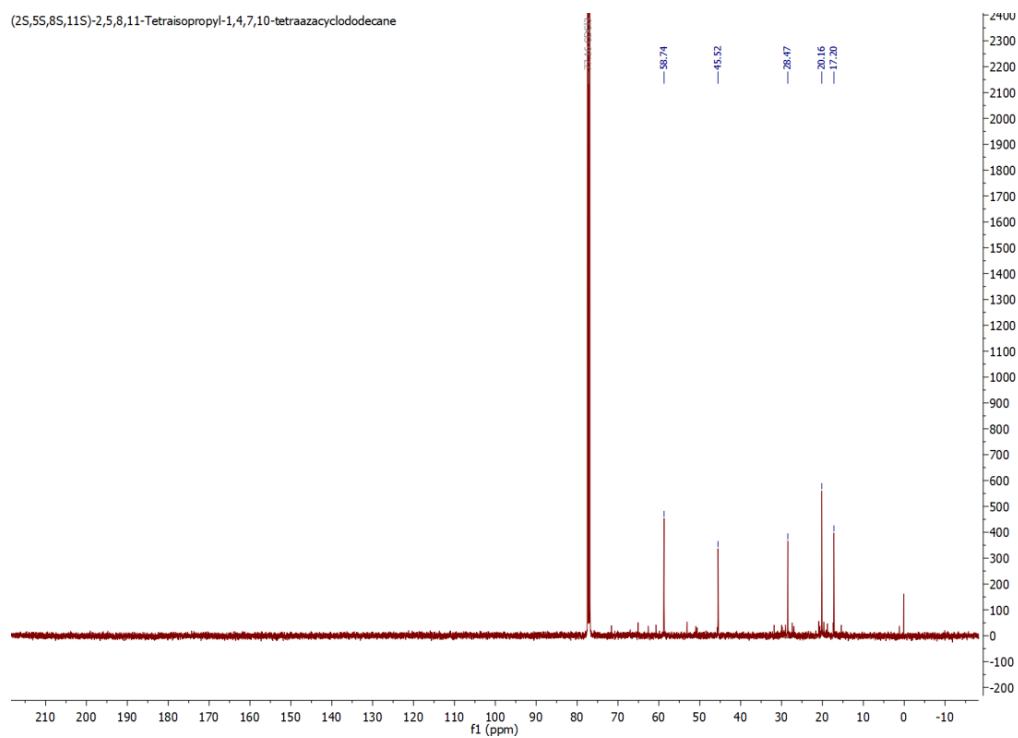


(2S,5S,8S,11S)-2,5,8,11-Tetraisopropyl-1,4,7,10-tetraazacyclododecane

¹H NMR

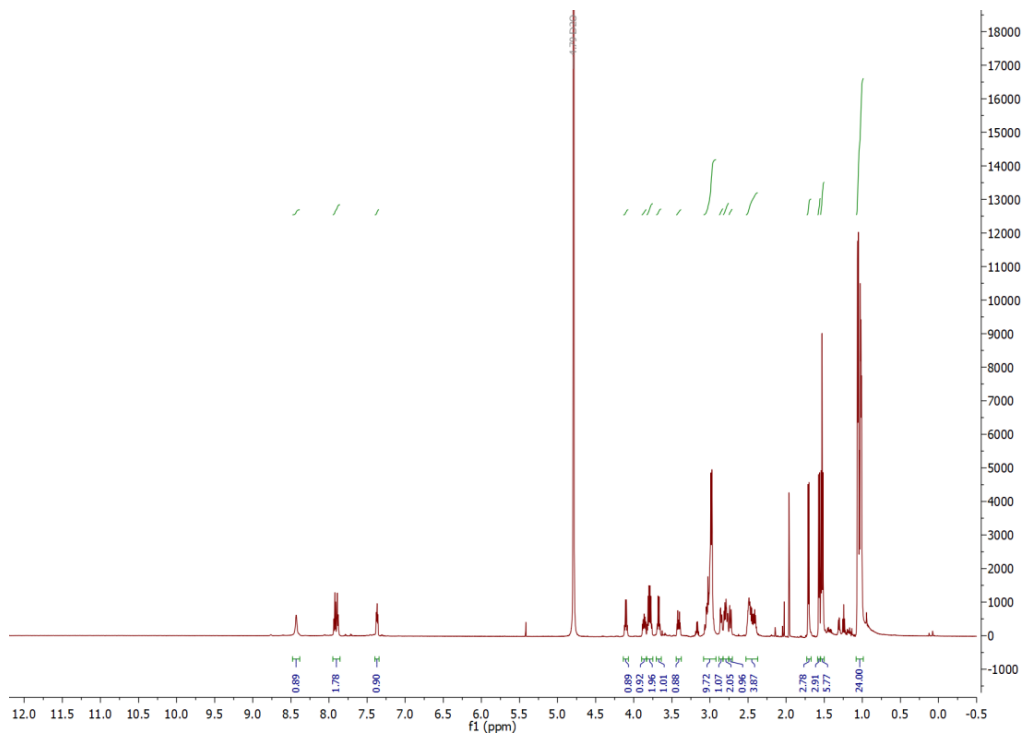


¹³C NMR

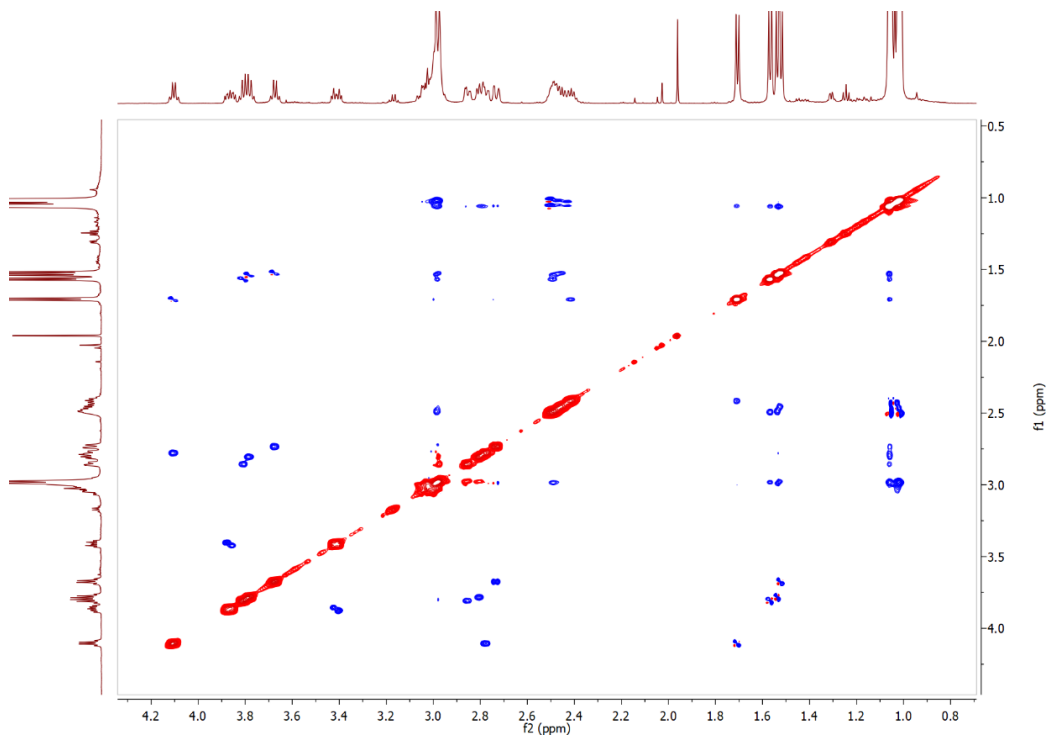


Lu-P4M4-DOTA

¹H NMR

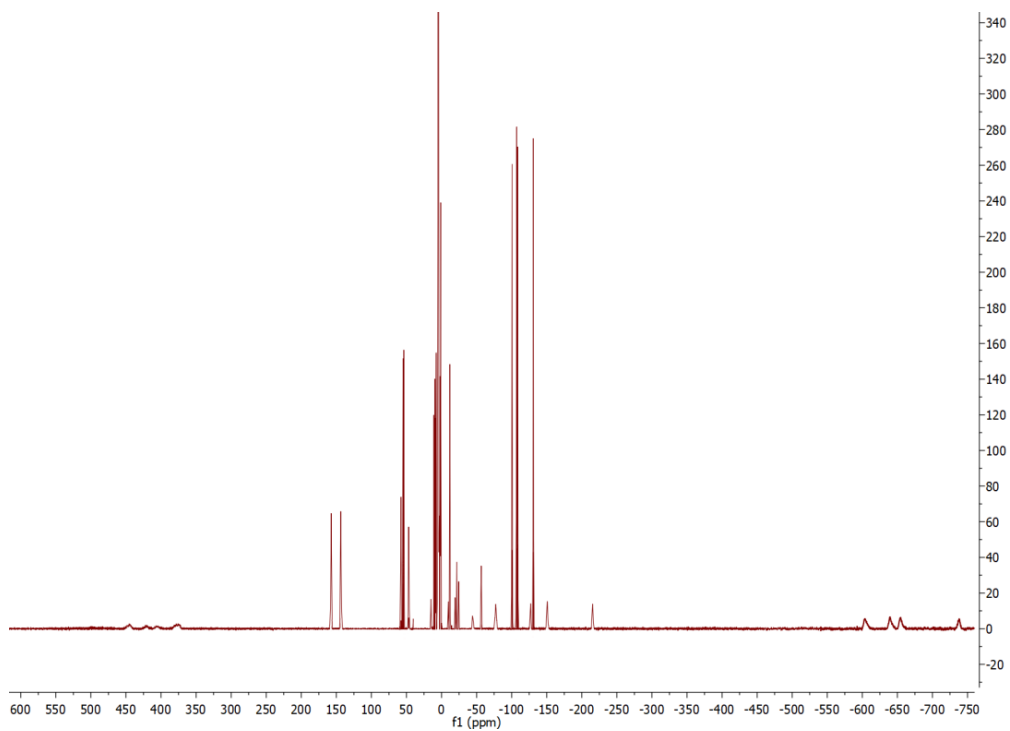


ROESY



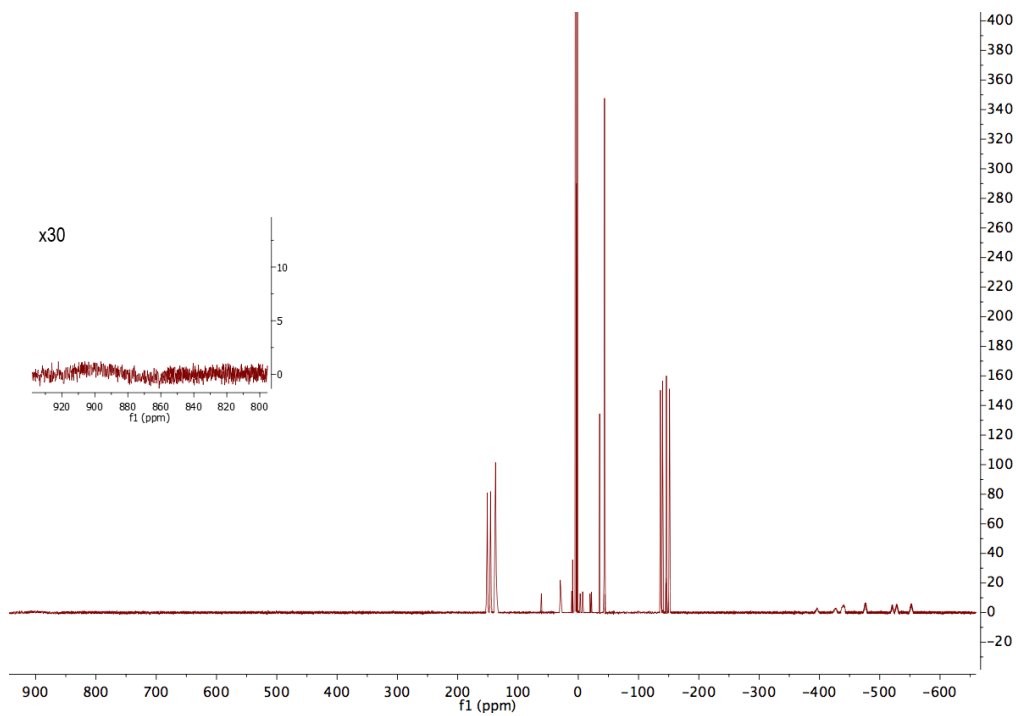
Dy-P4M4-DOTA

¹H NMR



Tm-P4M4-DOTA

¹H NMR



ESI-MS measurements of tagging reactions

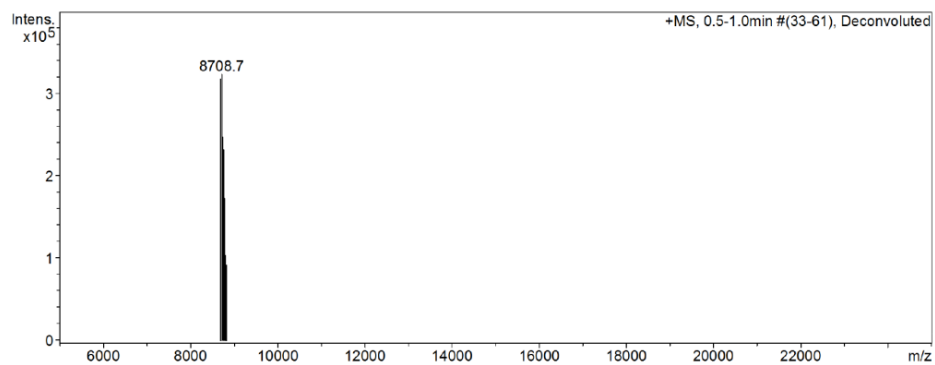


Figure S4: ESI-MS spectrum of ¹⁵N labelled ubiquitin S57C.

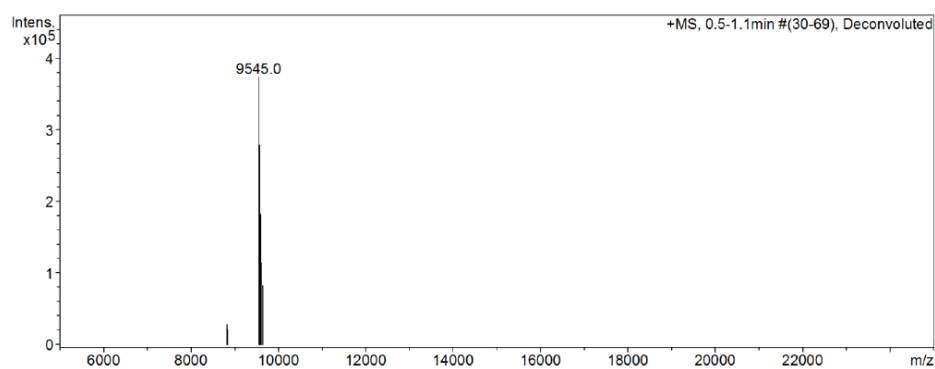


Figure S5: Confirmation of tagging reaction of ¹⁵N labelled ubiquitin S57C with Lu-P4M4-DOTA monitored by ESI-MS.

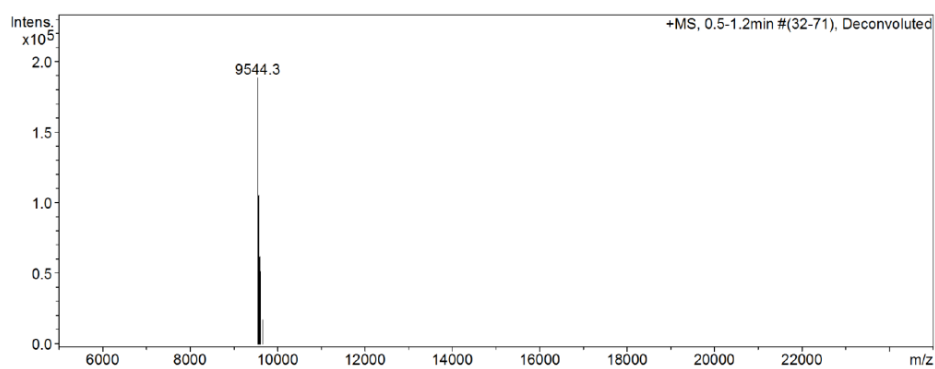


Figure S6: Confirmation of tagging reaction of ¹⁵N labelled ubiquitin S57C with Tm-P4M4-DOTA monitored by ESI-MS.

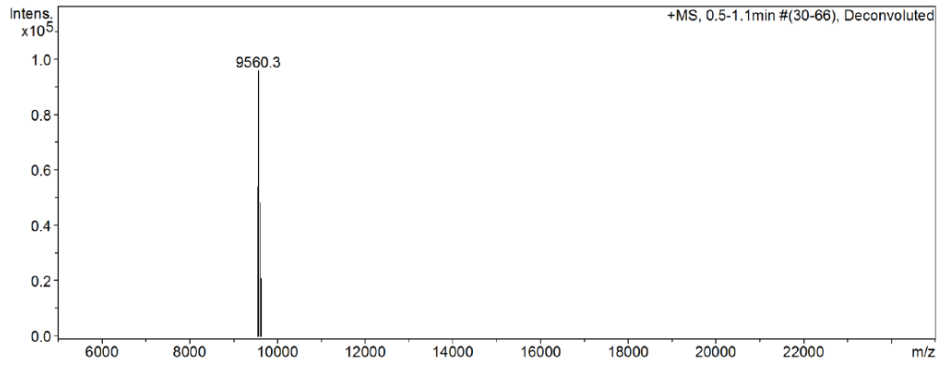


Figure S7: Confirmation of tagging reaction of ^{15}N labelled ubiquitin S57C with Dy-P4M4-DOTA monitored by ESI-MS.

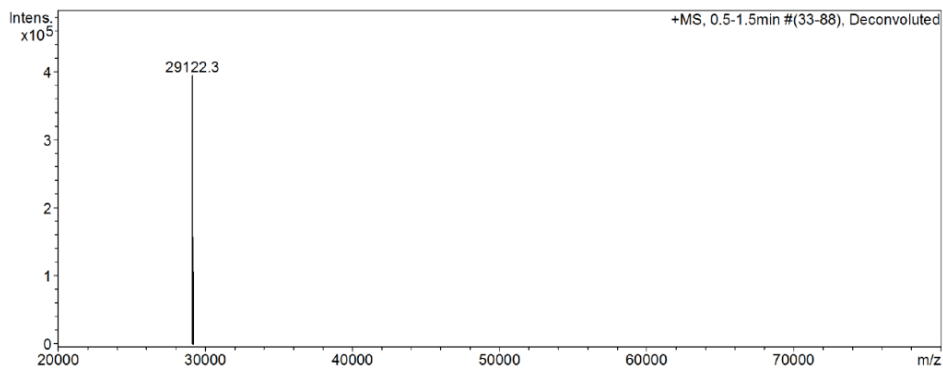


Figure S8: ESI-MS spectrum of selectively ^{15}N leucine labelled hCA II S166C.

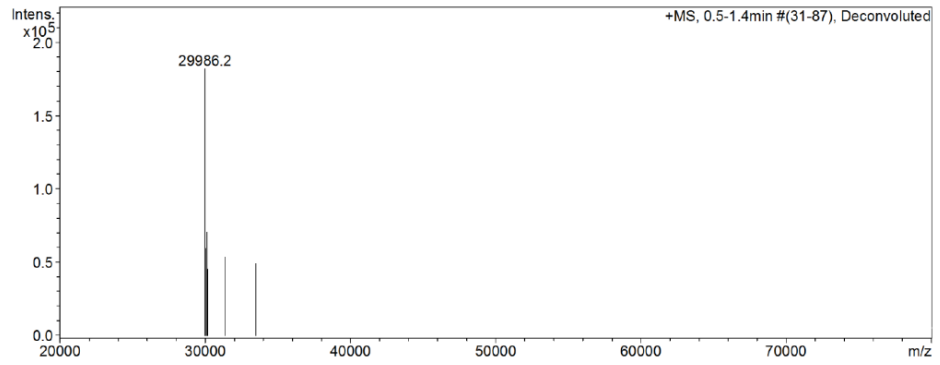


Figure S9: Confirmation of tagging reaction of selectively ^{15}N leucine labelled hCA II S166C with Lu-P4M4-DOTA monitored by ESI-MS.

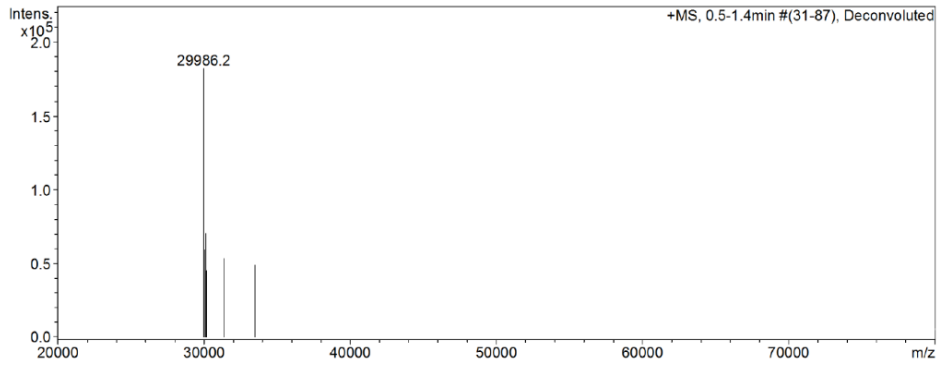


Figure S10: Confirmation of tagging reaction of selectively ¹⁵N leucine labelled hCA II S166C with Tm-P4M4-DOTA monitored by ESI-MS.

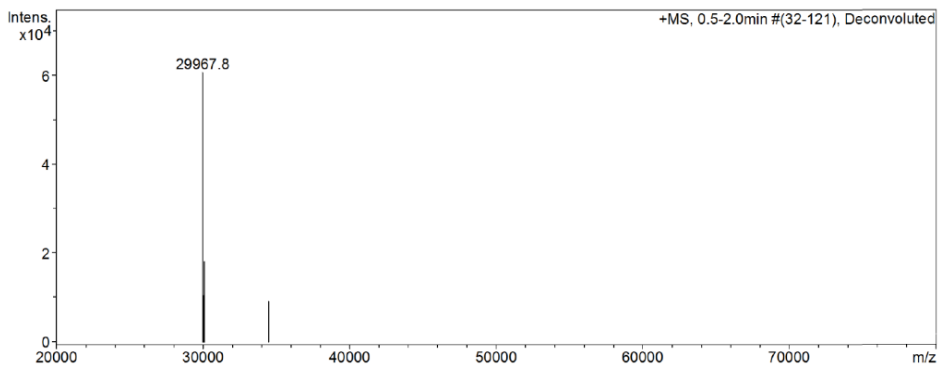


Figure S11: Confirmation of tagging reaction of selectively ¹⁵N leucine labelled hCA II S166C with Dy-P4M4-DOTA monitored by ESI-MS.

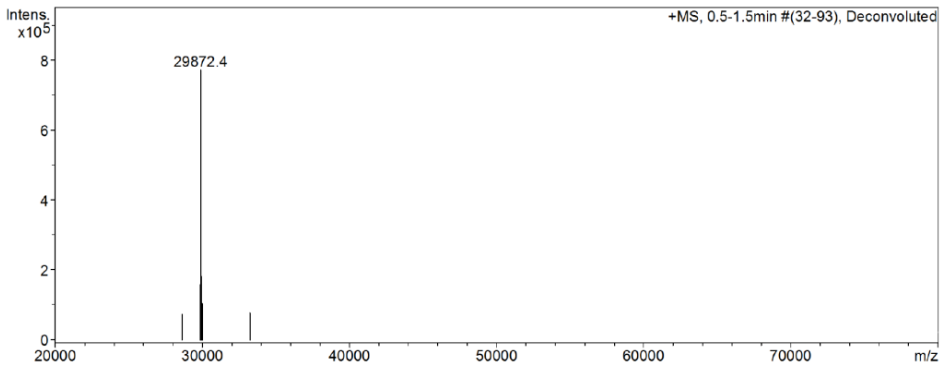


Figure S12: Confirmation of tagging reaction of selectively ¹⁵N leucine labelled hCA II S166C with Lu-DOTA-M8-(4R4S)-SSPy monitored by ESI-MS.

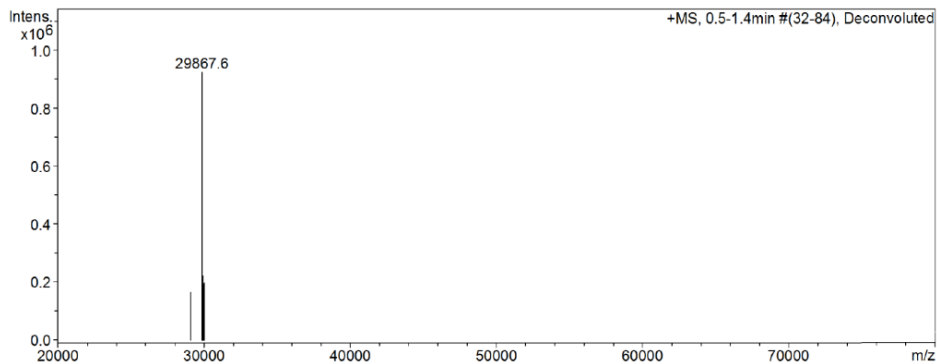


Figure S13: Confirmation of tagging reaction of selectively ¹⁵N leucine labelled hCA II S166C with Tm-DOTA-M8-(4R4S)-SSPy monitored by ESI-MS.

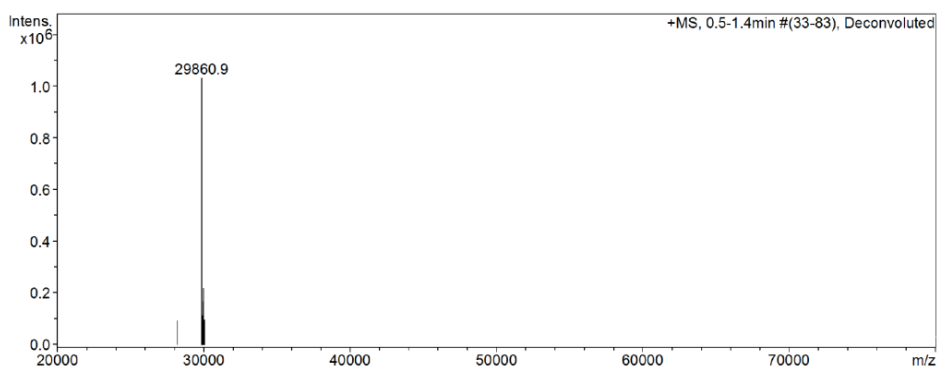


Figure S14: Confirmation of tagging reaction of selectively ¹⁵N leucine labelled hCA II S166C with Dy-DOTA-M8-(4R4S)-SSPy monitored by ESI-MS.

Shift list comparison of ^1H - ^{15}N HSQC spectra of Dy- and Lu-P4M4-DOTA-Ub^{S57C}

Table S1: Shift list comparison of ^1H - ^{15}N HSQC spectra of Dy- and Lu-P4M4-DOTA-Ub^{S57C}.

Residue	Reson. 1	Shift Dy	Shift Lu	Reson. 2	Shift Dy	Shift Lu
2Gln	H	9.80	8.89	N	123.54	122.71
3Ile	H	8.76	8.21	N	115.59	114.98
4Phe	H	9.06	8.54	N	119.06	118.59
5Val	H	9.50	9.22	N	121.49	121.22
6Lys	H	9.06	8.86	N	128.15	127.94
7Thr	H	8.81	8.68	N	115.72	115.59
12Thr	H	8.75	8.59	N	120.89	120.70
13Ile	H	9.69	9.48	N	127.93	127.77
14Thr	H	8.92	8.69	N	121.96	121.72
15Leu	H	9.04	8.68	N	125.50	125.23
16Glu	H	8.29	8.05	N	122.81	122.47
23Ile	H	8.67	8.43	N	121.50	121.22
25Asn	H	8.03	7.87	N	121.63	121.49
26Val	H	8.22	8.03	N	122.38	122.21
27Lys	H	8.63	8.49	N	119.17	119.00
28Ala	H	8.01	7.91	N	123.71	123.59
29Lys	H	7.90	7.78	N	120.41	120.30
30Ile	H	8.34	8.22	N	121.60	121.44
31Gln	H	8.56	8.48	N	123.80	123.62
32Asp	H	8.02	7.95	N	119.85	119.77
33Lys	H	7.43	7.33	N	115.57	115.47
34Glu	H	8.75	8.66	N	114.51	114.40
39Asp	H	8.52	8.46	N	113.74	113.73
40Gln	H	7.81	7.75	N	116.91	116.91
41Gln	H	7.48	7.42	N	118.25	118.21
42Arg	H	8.47	8.42	N	123.25	123.17
43Leu	H	8.81	8.77	N	124.57	124.48
44Ile	H	9.10	9.03	N	122.31	122.26
45Phe	H	8.71	8.77	N	125.02	125.02
48Lys	H	7.69	7.90	N	121.93	122.14
66Thr	H	9.12	8.66	N	117.92	117.47
67Leu	H	9.65	9.34	N	128.03	127.77
68His	H	9.26	9.14	N	119.59	119.43
69Leu	H	8.30	8.19	N	123.99	123.87
70Val	H	9.16	9.11	N	126.59	126.55
71Leu	H	8.08	8.05	N	123.05	123.03

Shift list comparison of ¹H-¹⁵N HSQC spectra of Tm- and Lu-P4M4-DOTA-Ub^{S57C}

Table S2: Shift list comparison of ¹H-¹⁵N HSQC spectra of Tm- and Lu-P4M4-DOTA-Ub^{S57C}.

Residue	Reson. 1	Shift Tm	Shift Lu	Reson. 2	Shift Tm	Shift Lu
2Gln	H	8.16	8.89	N	122.06	122.71
3Ile	H	7.57	8.21	N	114.27	114.98
4Phe	H	7.71	8.54	N	117.83	118.59
5Val	H	8.73	9.22	N	120.70	121.22
6Lys	H	8.40	8.86	N	127.48	127.94
7Thr	H	8.39	8.68	N	115.27	115.59
12Thr	H	8.33	8.59	N	120.40	120.70
13Ile	H	9.13	9.48	N	127.47	127.77
14Thr	H	8.38	8.69	N	121.35	121.72
15Leu	H	8.22	8.68	N	124.89	125.23
16Glu	H	7.85	8.05	N	122.25	122.47
23Ile	H	7.27	8.43	N	120.18	121.22
25Asn	H	7.47	7.87	N	121.13	121.49
27Lys	H	8.03	8.49	N	118.55	119.00
28Ala	H	7.64	7.91	N	123.32	123.59
29Lys	H	7.57	7.78	N	120.10	120.30
30Ile	H	7.94	8.22	N	121.16	121.44
31Gln	H	8.28	8.48	N	123.40	123.62
32Asp	H	7.82	7.95	N	119.63	119.77
33Lys	H	7.18	7.33	N	115.31	115.47
34Glu	H	8.49	8.66	N	114.20	114.40
39Asp	H	8.27	8.46	N	113.57	113.73
40Gln	H	7.54	7.75	N	116.76	116.91
41Gln	H	7.16	7.42	N	117.96	118.21
42Arg	H	8.10	8.42	N	122.78	123.17
43Leu	H	8.24	8.77	N	123.97	124.48
44Ile	H	8.48	9.03	N	121.67	122.26
45Phe	H	8.10	8.77	N	124.30	125.02
49Gln	H	8.18	8.59	N	122.76	123.21
59Tyr	H	1.04	7.15	N	109.83	115.48
66Thr	H	7.70	8.66	N	116.53	117.47
67Leu	H	8.59	9.34	N	127.01	127.77
69Leu	H	7.81	8.19	N	123.45	123.87
70Val	H	8.77	9.11	N	126.20	126.55
71Leu	H	7.84	8.05	N	122.81	123.03

Shift list comparison of ^1H - ^{15}N HSQC spectra of Dy- and Lu-P4M4-DOTA-SSPy attached to selectively ^{15}N leucine labelled hCA II S166C

Table S3: Shift list comparison of ^1H - ^{15}N HSQC spectra of Dy- and Lu-P4M4-DOTA-SSPy attached to selectively ^{15}N leucine labelled hCA II S166C.

Residue	Reson. 1	Shift Dy	Shift Lu	Reson. 2	Shift Dy	Shift Lu
44Leu	H	6.98	6.90	N	123.13	123.04
47Leu	H	9.01	8.90	N	126.25	126.16
57Leu	H	9.40	8.84	N	117.93	117.48
60Leu	H	11.05	8.62	N	126.48	123.79
79Leu	H	9.08	8.94	N	123.60	123.51
84Leu	H	7.81	7.70	N	120.43	120.30
90Leu	H	8.67	8.46	N	123.90	123.75
100Leu	H	7.59	7.27	N	119.32	119.01
118Leu	H	10.42	9.82	N	131.06	130.36
120Leu	H	9.45	8.94	N	124.19	123.53
141Leu	H	8.86	8.67	N	115.79	115.53
144Leu	H	8.74	8.38	N	128.52	128.18
148Leu	H	8.65	8.16	N	120.84	120.33
157Leu	H	8.53	7.47	N	117.93	116.75
184Leu	H	7.66	7.16	N	117.73	117.25
185Leu	H	6.94	6.56	N	113.54	113.10
198Leu	H	8.16	7.92	N	118.59	118.32
203Leu	H	9.34	9.08	N	120.74	120.47
204Leu	H	6.11	5.88	N	111.59	111.27
212Leu	H	9.19	8.90	N	126.88	126.60
224Leu	H	8.78	7.94	N	121.73	120.86
229Leu	H	7.99	7.02	N	120.02	119.14

Shift list comparison of ^1H - ^{15}N HSQC spectra of Tm- and Lu-P4M4-DOTA-SSPy attached to selectively ^{15}N leucine labelled hCA II S166C

Table S4: Shift list comparison of ^1H - ^{15}N HSQC spectra of Tm- and Lu-P4M4-DOTA-SSPy attached to selectively ^{15}N leucine labelled hCA II S166C.

Residue	Reson. 1	Shift Tm	Shift Lu	Reson. 2	Shift Tm	Shift Lu
44Leu	H	6.82	6.90	N	122.92	123.04
47Leu	H	8.80	8.90	N	126.06	126.16
57Leu	H	8.62	8.84	N	117.48	117.48
60Leu	H	5.65	8.62	N	120.36	123.79
79Leu	H	8.82	8.94	N	123.38	123.51
84Leu	H	7.59	7.70	N	120.21	120.30
90Leu	H	8.32	8.46	N	123.62	123.75
100Leu	H	6.73	7.27	N	118.50	119.01
118Leu	H	9.04	9.82	N	129.52	130.36
120Leu	H	8.36	8.94	N	122.90	123.53
141Leu	H	8.46	8.67	N	115.24	115.53
144Leu	H	7.96	8.38	N	127.74	128.18
148Leu	H	7.50	8.16	N	119.68	120.33
157Leu	H	6.25	7.47	N	115.40	116.75
184Leu	H	6.65	7.16	N	116.72	117.25
185Leu	H	6.16	6.56	N	112.57	113.10
198Leu	H	7.61	7.92	N	117.96	118.32
203Leu	H	8.73	9.08	N	120.02	120.47
204Leu	H	5.62	5.88	N	110.91	111.27
212Leu	H	8.55	8.90	N	126.33	126.60
224Leu	H	6.63	7.94	N	119.50	120.86
240Leu	H	8.32	8.59	N	125.54	125.72

Shift list comparison of ^1H - ^{15}N HSQC spectra of Dy- and Lu-DOTA-M8-(4R4S)-SSPy attached to selectively ^{15}N leucine labelled hCA II S166C

Table S5: Shift list comparison of ^1H - ^{15}N HSQC spectra of Dy- and Lu-DOTA-M8-(4R4S)-SSPy attached to selectively ^{15}N leucine labelled hCA II S166C.

Residue	Reson. 1	Shift Dy	Shift Lu	Reson. 2	Shift Dy	Shift Lu
44Leu	H	6.95	6.93	N	123.05	123.02
47Leu	H	8.93	8.93	N	126.16	126.15
57Leu	H	8.75	8.86	N	117.22	117.48
60Leu	H	10.11	8.64	N	125.50	123.79
79Leu	H	8.94	8.96	N	123.44	123.51
84Leu	H	7.75	7.72	N	120.36	120.31
90Leu	H	8.48	8.49	N	123.70	123.73
100Leu	H	7.46	7.29	N	119.14	118.99
118Leu	H	10.09	9.84	N	130.68	130.40
120Leu	H	9.16	8.96	N	123.85	123.52
141Leu	H	8.79	8.70	N	115.65	115.52
144Leu	H	8.52	8.40	N	128.26	128.14
148Leu	H	8.39	8.18	N	120.56	120.32
157Leu	H	7.76	7.50	N	117.03	116.73
184Leu	H	7.20	7.19	N	117.24	117.23
185Leu	H	6.61	6.59	N	113.16	113.09
189Leu	H	8.81	8.72	N	126.00	125.92
198Leu	H	8.10	7.96	N	118.43	118.27
203Leu	H	9.27	9.09	N	120.57	120.37
204Leu	H	6.05	5.89	N	111.41	111.27
212Leu	H	9.02	8.92	N	126.69	126.61
224Leu	H	8.36	7.97	N	121.27	120.87

Shift list comparison of ^1H - ^{15}N HSQC spectra of Tm- and Lu-DOTA-M8-(4R4S)-SSPy attached to selectively ^{15}N leucine labelled hCA II S166C

Table S6: Shift list comparison of ^1H - ^{15}N HSQC spectra of Tm- and Lu-DOTA-M8-(4R4S)-SSPy attached to selectively ^{15}N leucine labelled hCA II S166C.

Residue	Reson. 1	Shift Tm	Shift Lu	Reson. 2	Shift Tm	Shift Lu
44Leu	H	6.92	6.93	N	122.98	123.02
47Leu	H	8.95	8.93	N	126.15	126.15
57Leu	H	9.29	8.86	N	118.15	117.48
60Leu	H	6.35	8.64	N	121.13	123.79
79Leu	H	9.02	8.96	N	123.53	123.51
84Leu	H	7.72	7.72	N	120.26	120.31
90Leu	H	8.57	8.49	N	123.81	123.73
100Leu	H	6.94	7.29	N	118.63	118.99
118Leu	H	9.46	9.84	N	129.94	130.40
120Leu	H	8.72	8.96	N	123.31	123.52
141Leu	H	8.59	8.70	N	115.35	115.52
144Leu	H	8.24	8.40	N	127.97	128.14
148Leu	H	7.85	8.18	N	119.96	120.32
157Leu	H	7.30	7.50	N	116.50	116.73
184Leu	H	7.28	7.19	N	117.31	117.23
185Leu	H	6.64	6.59	N	113.04	113.09
189Leu	H	8.62	8.72	N	125.74	125.92
198Leu	H	7.75	7.96	N	118.01	118.27
203Leu	H	8.82	9.09	N	120.05	120.37
204Leu	H	5.67	5.89	N	111.02	111.27
212Leu	H	8.79	8.92	N	126.52	126.61
224Leu	H	7.19	7.97	N	120.01	120.87

Residual dipolar couplings measured in ^1H - ^{15}N HSQC IPAP spectra of Dy- and Lu-P4M4-DOTA-Ub^{S57C}

Table S7: Residual dipolar couplings measured in ^1H - ^{15}N HSQC IPAP spectra (900 MHz) of Dy- and Lu-P4M4-DOTA-Ub^{S57C}.

Residue	RDC (Hz)
27	-14.4
68	-6.7
39	11.7
48	4.6
23	0.4
28	-12.1
40	15.0
32	-7.7
45	-2.9
30	-7.6
34	-8.6
41	5.5
16	2.8
43	-9.5
4	4.8
29	-6.7
25	-6.3
2	-0.4
42	-13.1
67	15.3
69	-14.5
15	11.3
13	3.0
33	-0.7
44	-11.7
66	11.3
70	-12.4
7	-5.9
3	3.8
5	5.0
71	-0.3
14	0.3
31	-10.4
26	-1.4
6	-5.1
12	-9.4

Residual dipolar couplings measured in ^1H - ^{15}N HSQC IPAP spectra of Tm- and Lu-P4M4-DOTA-Ub^{S57C}

Table S8: Residual dipolar couplings measured in ^1H - ^{15}N HSQC IPAP spectra (900 MHz) of Tm- and Lu-P4M4-DOTA-Ub^{S57C}.

Residue	RDC (Hz)
27	12.7
39	-21.9
23	5.9
32	16.7
45	12.7
30	11.5
34	19.5
41	-2.0
49	-5.2
16	-4.9
43	20.0
4	-13.7
29	3.4
25	10.4
2	5.2
42	17.3
67	-13.9
14	-4.1
15	-16.1
13	-2.1
33	4.2
44	15.3
66	-20.8
70	15.0
7	10.0
3	-15.4
5	-6.7
71	2.2
69	17.6
31	18.9
6	4.2
12	9.2

Residual dipolar couplings measured in ^1H - ^{15}N HSQC IPAP spectra of Tm- and Lu-P4M4-DOTA-SSPy attached to selectively ^{15}N leucine labelled hCA II S166C

Table S9: Residual dipolar couplings measured in ^1H - ^{15}N HSQC IPAP spectra (600 MHz) of Tm- and Lu-P4M4-DOTA-SSPy attached to selectively ^{15}N leucine labelled hCA II S166C.

Residue	RDC (Hz)
141	11.3
118	-5.6
44	-7.8
224	-8.4
100	-2.6
144	14.1
57	26.2
198	-4.8
184	-11.7
84	1.3
148	-14.4
203	17.0
240	-19.3
120	4.7
157	-5.7
212	2.1
185	6.4

Correlation plots of experimental and back-calculated PCS and RDC

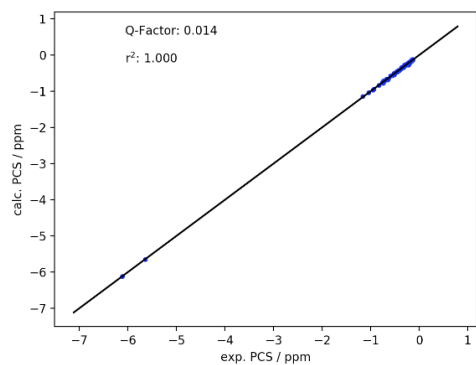


Figure S15: PCS correlation plot of ^{15}N labelled ubiquitin S57C labelled with Tm-P4M4-DOTA.

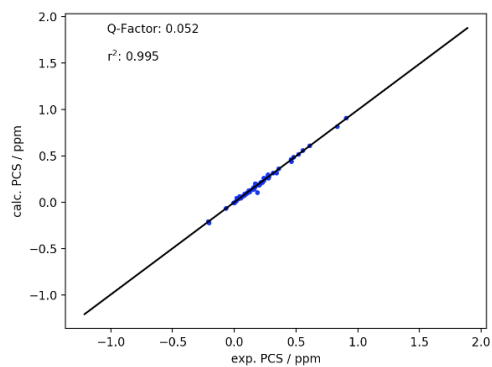


Figure S16: PCS correlation plot of ^{15}N labelled ubiquitin S57C labelled with Dy-P4M4-DOTA.

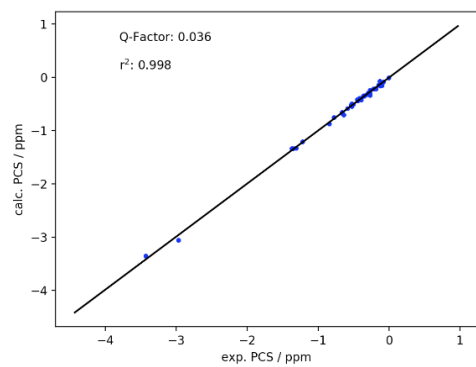


Figure S17: PCS correlation plot of selectively ^{15}N leucine labelled hCA S166C labelled with Tm-P4M4-DOTA.

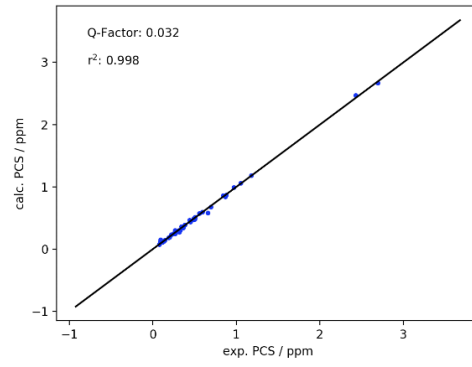


Figure S18: PCS correlation plot of selectively ^{15}N leucine labelled hCA S166C labelled with Dy-P4M4-DOTA.

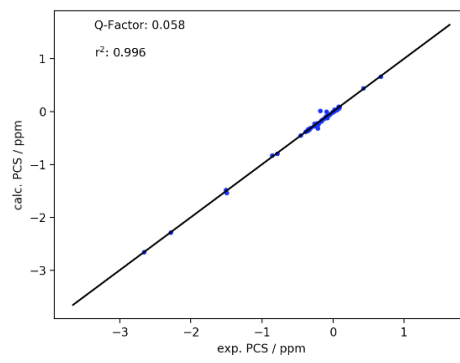


Figure S19: PCS correlation plot of selectively ^{15}N leucine labelled hCA S166C labelled with Tm-DOTA-M8-(4R4S)-SSPy.

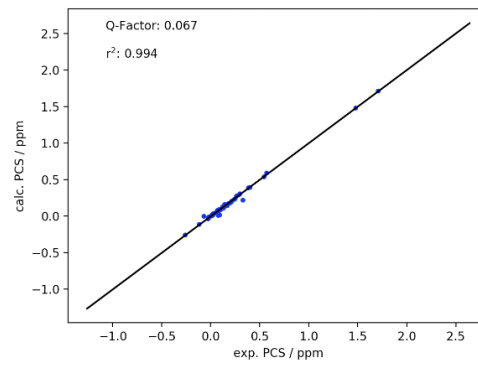


Figure S20: PCS correlation plot of selectively ^{15}N leucine labelled hCA S166C labelled with Dy-DOTA-M8-(4R4S)-SSPy.

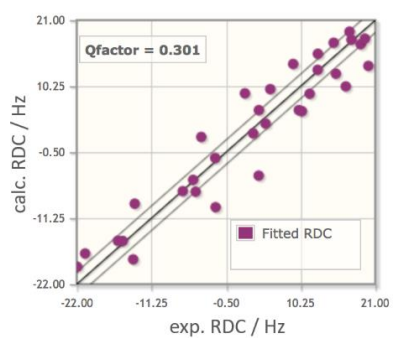


Figure S21: RDC correlation plot of ^{15}N labelled ubiquitin S57C labelled with Tm-P4M4-DOTA.

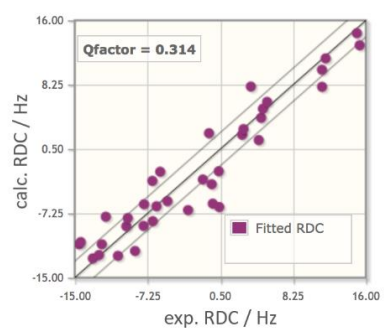


Figure S22: RDC correlation plot of ^{15}N labelled ubiquitin S57C labelled with Dy-P4M4-DOTA.

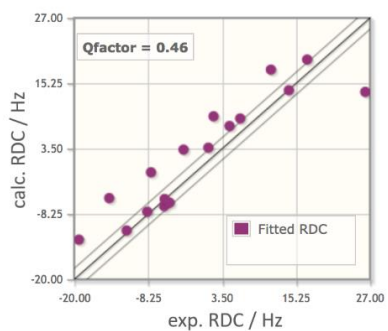


Figure S23: RDC correlation plot of selectively ^{15}N leucine labelled hCA S166C labelled with Tm-DOTA-M8-(4R4S)-SSPy.

Comparison of untagged ubiquitin S57C, Lu-DOTA-M8-(4R4S)-Ub^{S57C} and Lu-P4M4-DOTA-Ub^{S57C}

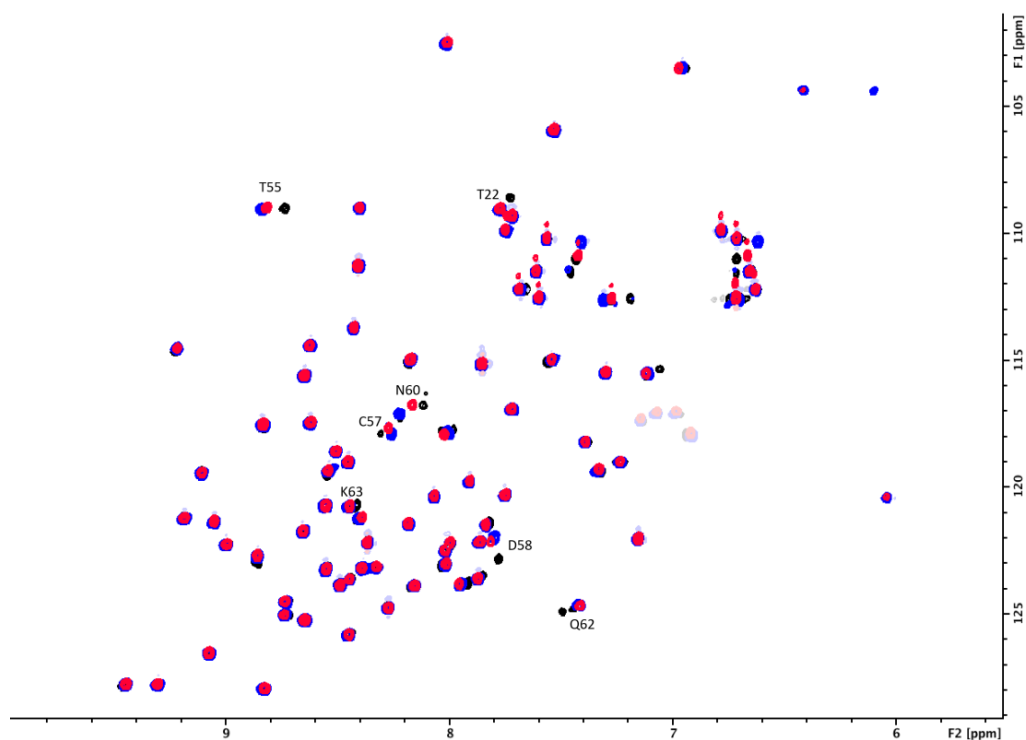


Figure S24: Overlay of ^1H - ^{15}N HSQC spectra of untagged ubiquitin S57C (pos. contours: black, neg. contours: grey), Lu-DOTA-M8-(4R4S)-Ub^{S57C} (pos. contours: red, neg. contours: fade red) and Lu-P4M4-DOTA-Ub^{S57C} (pos. contours: blue, neg. contours: fade blue) measured in 10 mM phosphate buffer with pH 6.0 at a temperature of 298 K on a 600 MHz Bruker Avance III NMR spectrometer equipped with a cryogenic QCI-F probe. Minor shift differences are observed for residues close to the tagging site. Negative peaks with a ^1H shift around 7.0 ppm and a ^{15}N shift around 117.5 ppm constitute aliased signals from the four arginine sidechains of ubiquitin S57C. Maximal shifts observed in ^1H - ^{15}N HSQC spectra of diamagnetic constructs were 0.10 ppm in ^1H dimension for residue T55 and 0.81 ppm in ^{15}N dimension for residue D58.

Comparison of donor atom positions for dysprosium- and thulium complexes of Ln-DOTA-M8-(4R4S)-SSPy and Ln-P4M4-DOTA obtained from DFT calculations

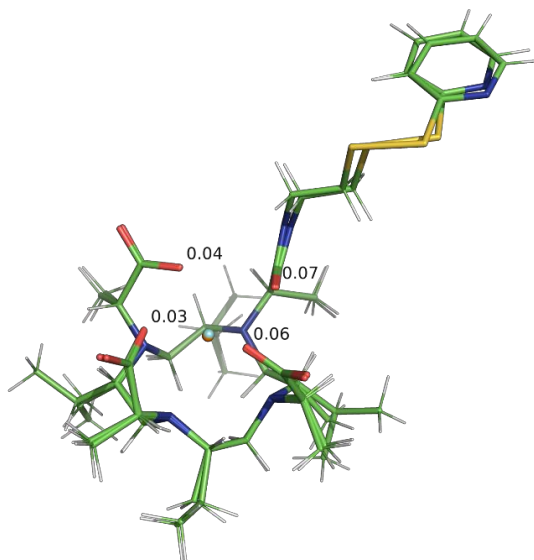


Figure S25: Overlay of the calculated DFT structures of Tm-DOTA-M8-(4R4S)-SSPy (Tm³⁺ ion: cyan) and Tm-P4M4-DOTA (Tm³⁺ ion: orange) depicting the minor differences in donor atom positions in Å. The structures were aligned by their ring core of the macrocycle using PyMOL.

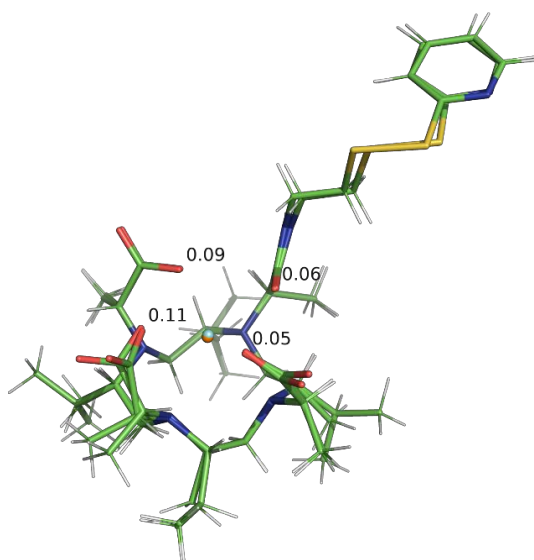


Figure S26: Overlay of the calculated DFT structures of Dy-DOTA-M8-(4R4S)-SSPy (Dy³⁺ ion: cyan) and Dy-P4M4-DOTA (Dy³⁺ ion: orange) depicting the minor differences in donor atom positions in Å. The structures were aligned by their ring core of the macrocycle using PyMOL.

Density functional theory calculations of lanthanide complexes – stabilization energies

Table S10: Stabilization energies for Lu-, Tm- and Dy-P4M4-DOTA obtained in DFT calculations.

Calculated energies of the different complexes M(L) (E_n / Hartree), L = P4M4-DOTA		Stabilization (kJ/mol)
Lu(L) TSAP	Lu(L) SAP	
-17028.45263	-17028.45754	-12.9
Lu(L) TSAP solv.	Lu(L) SAP solv.	
-17028.53926	-17028.55167	-32.6
Tm(L) TSAP	Tm(L) SAP	
-16117.43721	-16117.44756	-27.2
Tm(L) TSAP solv.	Tm(L) SAP solv.	
-16117.52065	-16117.53077	-26.6
Dy(L) TSAP	Dy(L) SAP	
-14819.35856	-14819.36828	-25.5
Dy(L) TSAP solv.	Dy(L) SAP solv.	
-14819.44106	-14819.45122	-26.7

References

- [1] M. Frigerio, M. Santagostino, S. Sputore, *The Journal of Organic Chemistry* **1999**, *64*, 4537-4538.

3.3 P4T-DOTA - A lanthanide chelating tag combining a highly sterically overcrowded backbone with a reductively stable linker

Reproduced from D. Joss, D. Häussinger, *Chemical Communications*, **2019**, 55, 10543-10546 with permission from the Royal Society of Chemistry.



Cite this: *Chem. Commun.*, 2019, 55, 10543

Received 18th June 2019,
Accepted 8th August 2019

DOI: 10.1039/c9cc04676c

rsc.li/chemcomm

P4T-DOTA – a lanthanide chelating tag combining a sterically highly overcrowded backbone with a reductively stable linker†

Daniel Joss  and Daniel Häussinger *

Herein we report a DOTA-based lanthanide chelating tag (LCT) with rigidified backbone and a reduction-stable linker. The newly developed tag induces strong pseudocontact shifts suitable for paramagnetic protein nuclear magnetic resonance spectroscopy and the obtained anisotropic susceptibility parameters are in the range of the best performing LCTs.

Pseudocontact shifts (PCS) and residual dipolar couplings (RDC) obtained by using lanthanide chelating tags (LCT) yield valuable restraints for investigating protein structures, dynamics and interactions in solution.^{1–19} The stereo-specifically methyl substituted 1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid (DOTA)-based chelators provide a sufficiently rigidified scaffold for observation of significant structural restraints.^{9,20,21} Interestingly, the methyl substituents on the basic macrocyclic scaffold adopt an equatorial-upper position in the final lanthanide complex.^{22,23} The methyl substituents thereby arrange in the most suitable way to provide a favourable cavity for the lanthanide ion and minimize the steric repulsion between each other. Furthermore, a crucial factor in the design of the LCTs is to obtain a lanthanide complex that shows only one diastereomer and only one conformation of the pendant arms coordinating to the lanthanide ion.¹⁷ The conformationally locked and single diastereomeric LCT provides then only one signal set in ¹H–¹⁵N HSQC experiments and yields strong paramagnetic effects due to its rigidity and immobilization on the proteins surface.¹⁷ As shown by Joss *et al.*, introduction of even more bulkier isopropyl substituents on the backbone of the LCT can significantly enhance the tensor parameters when compared to its methyl substituted predecessor.²⁴ In order to circumvent the inherent instability of disulphide linkers towards a reductive environment, various pyridinesulphone- and iodoacetamide linkers have recently

been developed. Therefore, we envisioned to synthesize a lanthanide chelating tag offering the combination of a rigidified backbone and a reductively-stable linker (Fig. 1), that enables convenient and fast protein tagging. The resulting thioether linkage is stable under reductive conditions and has been demonstrated to allow for observation of PCS and RDC in intact eukaryotic cells.^{2,3}

In order to synthesize the newly designed tags, we combined the synthetic approaches by Joss *et al.* (isopropyl-substituted backbone)²⁴ and Müntener *et al.* (thiazolo linker)¹⁸ and developed them further to yield the target molecule.

To explore the performance, *i.e.* the range of PCS and RDC as well as the associated $\Delta\chi$ parameters of the Ln-P4T-DOTA tag (P4T-DOTA: (2*R*,2'*R*,2''*R*)-2,2',2''-((2*S*,5*S*,8*S*,11*S*)-2,5,8,11-tetraisopropyl-10-((2-(methylsulphonyl)thiazolo[5,4-*b*]pyridine-5-yl)methyl)-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)tripropionate), the Dy-, Tm- and Lu complexes of P4T-DOTA were synthesized and conjugated to ubiquitin S57C, ubiquitin K48C and to selectively ¹⁵N leucine labelled human carbonic anhydrase S166C. The newly developed LCT was benchmarked by analysing PCS and RDC (Table 1).

P4T-DOTA delivers large PCS on all tested protein constructs exceeding most current high-performance lanthanide chelating tags (Fig. 2, 3 and 5).^{3,9,11,15,17,18,28,29}

The obtained tensor shapes resemble to the ones found for Ln-DOTA-M8-(4*R*4*S*)-SSPy and Ln-P4M4-DOTA, *i.e.* a significantly less rhombic tensor for the thulium complex when compared to

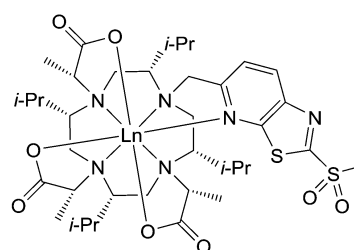


Fig. 1 Structure of Ln-P4T-DOTA in 1(δδδδ) conformation.

Department of Chemistry, University of Basel, St. Johannis-Ring 19, 4056 Basel, Switzerland. E-mail: daniel.haessinger@unibas.ch

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c9cc04676c

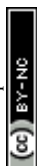


Table 1 Properties of the induced axial and rhombic components of the paramagnetic susceptibility tensors ($\Delta\chi_{ax}$ and $\Delta\chi_{rh}$), metal position in PDB coordinate frame (X_{metal} , Y_{metal} , Z_{metal}), Euler angles (α , β , γ) and quality factor (Q , mathematical definition given in ESI) on ubiquitin S57C (pH 6.0), ubiquitin K48C (pH 6.0) and hCA II S166C (pH 6.8) at 298 K

Protein mutant	PDB	No. PCS	Ln ³⁺	$\Delta\chi_{ax}$ (10^{-32} m ³)	$\Delta\chi_{rh}$ (10^{-32} m ³)	X_{metal} (Å)	Y_{metal} (Å)	Z_{metal} (Å)	α (°)	β (°)	γ (°)	Q (%)
Ubiquitin S57C	1UBI ²⁵	50	Dy	54.0	27.0	21.6	14.5	6.0	151.7	85.5	132.4	4.5
		72	Tm	39.3	14.6	21.6	14.5	6.0	60.0	34.7	2.9	4.6
Ubiquitin K48C		32	Dy	-53.7	-23.5	20.5	19.9	25.8	117.3	117.4	118.5	21.5
		54	Tm	39.6	13.4	20.5	19.9	25.8	91.4	106.5	97.3	3.6
hCA II S166C	3KS3 ²⁶	40	Dy	-46.1	-30.5	-15.9	-3.9	-10.8	24.4	43.7	38.9	11.4
		46	Tm	44.3	4.8	-15.9	-3.9	-10.8	174.7	157.0	29.0	5.8

Protein mutant	PDB	No. RDC	Ln ³⁺	$\Delta\chi_{ax}$ (10^{-32} m ³)	$\Delta\chi_{rh}$ (10^{-32} m ³)	X_{metal} (Å)	Y_{metal} (Å)	Z_{metal} (Å)	α (°)	β (°)	γ (°)	Q (%)
Ubiquitin S57C	2MJB ²⁷	24	Dy	-45.5	-26.2	15.5	-9.0	4.5	79.2	33.4	156.9	15.1
		36	Tm	-36.3	-12.3	15.5	-9.0	4.5	40.1	90.7	168.8	23.2

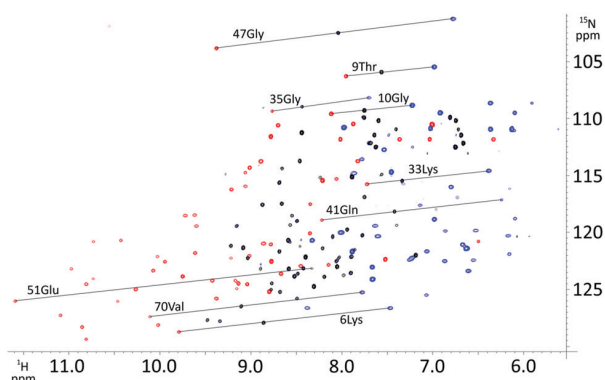


Fig. 2 Overlay of ¹H-¹⁵N HSQC spectra of Dy- (blue), Tm- (red), and Lu-P4T-Ub^{557C} (black).

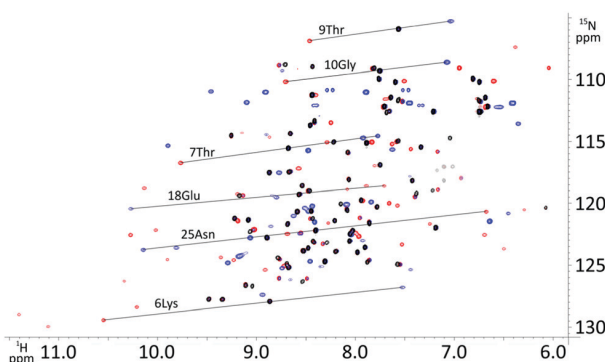


Fig. 3 Overlay of ¹H-¹⁵N HSQC spectra of Dy- (blue), Tm- (red), and Lu-P4T-Ub^{K48C} (black).

the tensor of the dysprosium complexes (Fig. 4, 6 and 7). The tensors for Tm-P4T-Ub^{S57C/K48C} as well as for Tm-P4T attached to selectively ¹⁵N leucine labelled hCA II S166C display a much more favourable motional averaging in comparison to Tm-DOTA-M8-(4R4S)-SSPy, and ensure in this way that the magnetic anisotropy of the lanthanide is efficiently transferred to the protein. This feature can be attributed to the rigid and very short thiazolo-linker, that only enables rotation around the C_{thiazolo}-S_{Cys} bond. Furthermore an orientation of the tag is enforced, so that the large

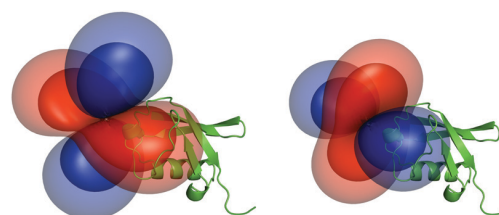


Fig. 4 Tensors generated by the dysprosium (left) and thulium (right) complex and their relative orientation to ubiquitin S57C (PCS isosurfaces set to 1.5 ppm (outer layer) and 4.0 ppm (inner layer)).

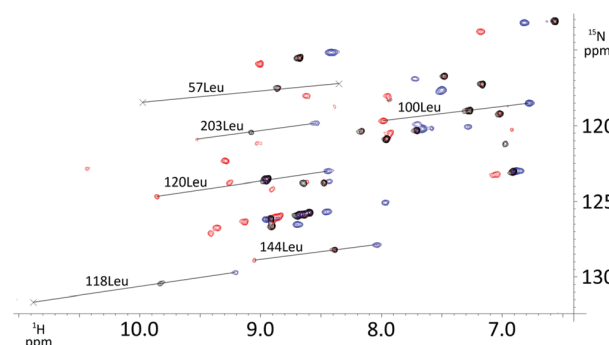


Fig. 5 Overlay of ¹H-¹⁵N HSQC spectra of Dy- (blue), Tm- (red), and Lu-P4T (black) attached to selectively ¹⁵N leucine labelled human carbonic anhydrase II S166C.

axial lobe of the isosurfaces is colinear with the C_{thiazolo}-S_{Cys} bond and therefore, less diminished by rotational averaging (see Fig. 4 and 6–8).¹⁸

When compared to the methyl-substituted thiazolo tag described recently by Müntener *et al.*,¹⁸ the isopropyl-substituted thiazolo tag shows an increase in tensor magnitudes due to the sterically more crowded ligand (Table 1).

Besides the favourable $\Delta\chi$ -tensor properties, the reductively stable linker offers new possibilities, as *e.g.* applications in in-cell NMR, when compared to the Ln-P4M4-DOTA tag, that can only be employed under non-reductive conditions. The strong paramagnetic relaxation enhancement (PRE), *e.g.* generated for the Dy-P4T-Ub^{K48C} construct leads to relatively few detectable signals. However, the results obtained for the



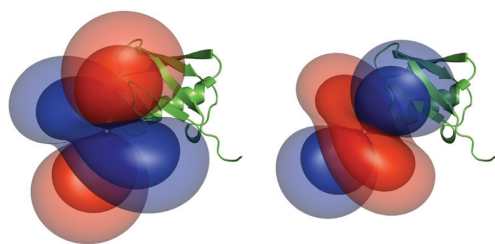


Fig. 6 Tensors generated by the dysprosium (left) and thulium (right) complex and their relative orientation to ubiquitin K48C (PCS isosurfaces set to 1.5 ppm (outer layer) and 4.0 ppm (inner layer)).

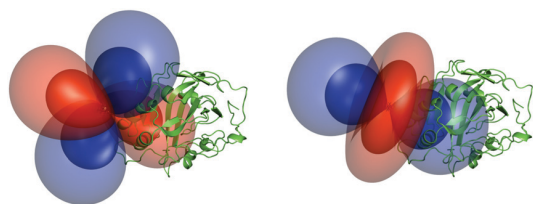


Fig. 7 Tensors generated by the dysprosium (left) and thulium (right) complex and their relative orientation to selectively ^{15}N leucine labelled hCA II (PCS isosurfaces set to 0.5 ppm (outer layer) and 2.0 ppm (inner layer)).

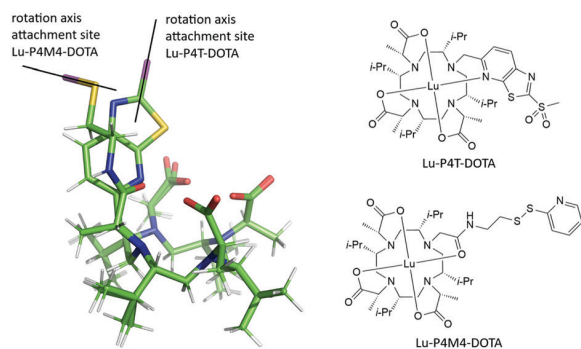


Fig. 8 Overlay of DFT structures of Lu-P4T and Lu-P4M4 and their structures. Lu-P4T (top right), Lu-P4M4 (lower right), attachment point of the protein's cysteine residue (magenta).

thulium tag when attached to ubiquitin S57C and hCA II S166C, *i.e.* the very large pseudocontact shifts combined with moderate PRE, renders the thulium P4T tag as an ideal solution to investigate not only small and medium sized proteins up to 29 kDa but shows potential for future applications to proteins with significantly larger size.

In order to further characterize the LCT presented in this study, geometries of the Lu-, Tm- and Dy complexes were calculated using density functional theory (DFT) calculations. For the calculations, BP86 was used as functional, SARC-TZVP as basis set for the ligands, while SARC2-QZVP was used as basis set for the lanthanide metal. The calculations were performed using the relativistic ZORA approximation, as well as the RI approximation to speed up the calculations. To model the water solvent, CPCM solvent model was implemented into the calculations. The corresponding references are included in the SI.

Table 2 Stabilization energies obtained *in vacuo* and water solvent (indicated as solv.). TSAP = twisted square antiprism, SAP = square antiprism

Calculated energies of the different complexes Ln(L) given in Hartree, L = P4T-DOTA	Stabilization $\Delta(\delta\delta\delta\delta)$ (SAP) over $\Delta(\delta\delta\delta\delta)$ (TSAP) (kJ mol^{-1})	
Lu(L) TSAP −17023.78390	Lu(L) SAP −17023.78679	−7.6
Lu(L) TSAP solv. −17023.87054	Lu(L) SAP solv. −17023.87706	−17.1
Tm(L) TSAP −16112.77041	Tm(L) SAP −16112.77262	−5.8
Tm(L) TSAP solv. −16112.85311	Tm(L) SAP solv. −16112.85716	−10.6
Dy(L) TSAP −14814.69023	Dy(L) SAP −14814.69733	−18.7
Dy(L) TSAP solv. −14814.77349	Dy(L) SAP solv. −14814.78084	−19.3

Ln-DOTA-M8-(8*S*)-SSPy complexes show two conformational isomers depending on the ionic radius of the coordinated lanthanide, whereas the (4*R*4*S*) stereoisomer shows exclusively a $\Delta(\delta\delta\delta\delta)$ conformation.^{17,22} Based on the performed calculations with and without implicit solvent model, a clear stabilization of all investigated complexes towards a $\Delta(\delta\delta\delta\delta)$ geometry is observed, a result that matches the outcomes for related tags (Table 2).^{17,18} The obtained stabilization energies for the square antiprism (SAP) conformation of the lanthanide complexes in an implicit water solvent of 17.1 (lutetium), 10.6 (thulium) and 19.3 kJ mol^{-1} (dysprosium) correspond to equilibrium constants of 999, 73 and 2388 towards the favoured SAP conformation. Due to the higher steric demand of the thiazolo ligand close to the ninth coordination site when compared to the amide ligand of DOTA-M8-(4*R*4*S*)-SSPy and the experimental evidence by Strickland *et al.* that for the Yb-DOTA-M8-(8*S*)-SSPy with a significantly less hindered apical coordination site there is no coordination of a water molecule,³⁰ no calculations with an explicit water molecule on the ninth coordination site were performed.

Interestingly, from an overlay of the DFT structures of Lu-P4M4 and Lu-P4T, the angle of the LCT to the protein can be estimated (Fig. 8). Two striking differences can be observed: (i) the linker in Ln-P4T, which is rigidified by the non-flexible aromatic system, is significantly shorter and more rigid than the corresponding linker in Ln-P4M4, (ii) while Ln-P4T is attached in a favourable angle to the protein in terms of motional averaging, Ln-P4M4 is more prone to averaging of the magnetic anisotropy.

To conclude, a new, strongly paramagnetic lanthanide chelating tag is presented that yields pseudocontact shifts in a very high range, exhibits large anisotropy tensors for both employed lanthanide ions and forms a reductively stable linkage to the target protein. The newly developed tag was benchmarked on three different protein constructs, ubiquitin S57C, ubiquitin K48C and hCA II S166C. When compared to its predecessors, the presented LCT yields strongly enhanced pseudocontact shifts due to the very rigid and short linker in combination with a highly sterically crowded backbone. In order to enable further studies on large proteins, protein



complexes and other biomacromolecules by PCS NMR spectroscopy, the development of high-performance LCTs will be continued.

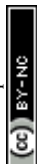
The Chemistry Department of the University of Basel and the Swiss National Science Foundation grant 200021_130263 are acknowledged for financial support. Biological structures were generated using the open source software PyMOL (<http://www.pymol.org/>). Calculations were performed at sciCORE (<http://scicore.unibas.ch/>) scientific computing core facility at University of Basel. C. E. Housecroft, E. C. Constable, T. Müntener and R. Vogel are acknowledged for helpful discussions.

Conflicts of interest

There are no conflicts of interest to declare.

Notes and references

- C. Nitsche and G. Otting, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2017, **98–99**, 20–49.
- B.-B. Pan, F. Yang, Y. Ye, Q. Wu, C. Li, T. Huber and X.-C. Su, *Chem. Commun.*, 2016, **52**, 10237–10240.
- T. Müntener, D. Häussinger, P. Selenko and F.-X. Theillet, *J. Phys. Chem. Lett.*, 2016, **7**, 2821–2825.
- Y. Hikone, G. Hirai, M. Mishima, K. Inomata, T. Ikeya, S. Arai, M. Shirakawa, M. Sodeoka and Y. Ito, *J. Biomol. NMR*, 2016, **66**, 99–110.
- W.-M. Liu, M. Overhand and M. Ubbink, *Coord. Chem. Rev.*, 2014, **273–274**, 2–12.
- C. T. Loh, K. Ozawa, K. L. Tuck, N. Barlow, T. Huber, G. Otting and B. Graham, *Bioconjugate Chem.*, 2013, **24**, 260–268.
- G. Otting, *Annu. Rev. Biophys.*, 2010, **39**, 387–405.
- X.-C. Su, H. Liang, K. V. Loscha and G. Otting, *J. Am. Chem. Soc.*, 2009, **131**, 10352–10353.
- D. Häussinger, J.-R. Huang and S. Grzesiek, *J. Am. Chem. Soc.*, 2009, **131**, 14761–14767.
- X.-C. Su, B. Man, S. Beeren, H. Liang, S. Simonsen, C. Schmitz, T. Huber, B. A. Messerle and G. Otting, *J. Am. Chem. Soc.*, 2008, **130**, 10486–10487.
- P. H. J. Keizers, A. Saragliadis, Y. Hiruma, M. Overhand and M. Ubbink, *J. Am. Chem. Soc.*, 2008, **130**, 14802–14812.
- G. Pintacuda, A. Y. Park, M. A. Keniry, N. E. Dixon and G. Otting, *J. Am. Chem. Soc.*, 2006, **128**, 3696–3702.
- K. D. Brewer, T. Bacaj, A. Cavalli, C. Camilloni, J. D. Swarbrick, J. Liu, A. Zhou, P. Zhou, N. Barlow, J. Xu, A. B. Seven, E. A. Prinslow, R. Voleti, D. Häussinger, A. M. J. J. Bonvin, D. R. Tomchick, M. Vendruscolo, B. Graham, T. C. Südhof and J. Rizo, *Nat. Struct. Mol. Biol.*, 2015, **22**, 555.
- V. Gaponenko, A. S. Altieri, J. Li and R. A. Byrd, *J. Biomol. NMR*, 2002, **24**, 143–148.
- F. Peters, M. Maestre-Martinez, A. Leonov, L. Kovacic, S. Becker, R. Boelens and C. Griesinger, *J. Biomol. NMR*, 2011, **51**, 329–337.
- J. Wöhnert, K. J. Franz, M. Nitz, B. Imperiali and H. Schwalbe, *J. Am. Chem. Soc.*, 2003, **125**, 13338–13339.
- D. Joss, R. M. Walliser, K. Zimmermann and D. Häussinger, *J. Biomol. NMR*, 2018, 29–38.
- T. Müntener, J. Kottelat, A. Huber and D. Häussinger, *Bioconjugate Chem.*, 2018, **29**, 3344–3351.
- K. Zimmermann, D. Joss, T. Müntener, E. S. Nogueira, L. Knörr, M. Schäfer, F. W. Monnard and D. Häussinger, *Chem. Sci.*, 2019, **10**, 5064–5072.
- R. S. Ranganathan, N. Raju, H. Fan, X. Zhang, M. F. Tweedle, J. F. Desreux and V. Jacques, *Inorg. Chem.*, 2002, **41**, 6856–6866.
- R. S. Ranganathan, R. K. Pillai, N. Raju, H. Fan, H. Nguyen, M. F. Tweedle, J. F. Desreux and V. Jacques, *Inorg. Chem.*, 2002, **41**, 6846–6855.
- A. C. L. Opina, M. Strickland, Y.-S. Lee, N. Tjandra, R. A. Byrd, R. E. Swenson and O. Vasalatiy, *Dalton Trans.*, 2016, **45**, 4673–4687.
- D. Parker, R. S. Dickins, H. Puschmann, C. Crossland and J. A. K. Howard, *Chem. Rev.*, 2002, **102**, 1977–2010.
- D. Joss, M.-S. Bertrams and D. Häussinger, *Chem. – Eur. J.*, 2019, DOI: 10.1002/chem.201901692.
- R. Ramage, J. Green, T. W. Muir, O. M. Ogunjobi, S. Love and K. Shaw, *Biochem. J.*, 1994, **299**(pt 1), 151–158.
- B. S. Avvaru, C. U. Kim, K. H. Sippel, S. M. Gruner, M. Agbandje-McKenna, D. N. Silverman and R. McKenna, *Biochemistry*, 2010, **49**, 249–251.
- A. S. Maltsev, A. Grishaev, J. Roche, M. Zasloff and A. Bax, *J. Am. Chem. Soc.*, 2014, **136**, 3752–3755.
- B. Graham, C. T. Loh, J. D. Swarbrick, P. Ung, J. Shin, H. Yagi, X. Jia, S. Chhabra, N. Barlow, G. Pintacuda, T. Huber and G. Otting, *Bioconjugate Chem.*, 2011, **22**, 2118–2125.
- F. Yang, X. Wang, B.-B. Pan and X.-C. Su, *Chem. Commun.*, 2016, **52**, 11535–11538.
- M. Strickland, C. D. Schwieters, C. Göbl, A. C. L. Opina, M.-P. Strub, R. E. Swenson, O. Vasalatiy and N. Tjandra, *J. Biomol. NMR*, 2016, **66**, 125–139.



Supporting Information

P4T-DOTA – A lanthanide chelating tag combining a sterically highly overcrowded backbone with a reductively stable linker

Daniel Joss^a and Daniel Häussinger^{*,a}

^a Department of Chemistry, University of Basel, St. Johanns-Ring 19, CH-4056 Basel, daniel.haeussinger@unibas.ch

General Remarks

Unless otherwise stated, reactions were performed under an argon atmosphere and chemicals were used as received without further purification. Ubiquitin S57C and K48C were expressed as described previously by Sass et al., selectively ¹⁵N leucine labelled hCA II S166C as described by Varghese et al.^{1,2} Tagging reactions were performed with a conversion of > 95% at a protein concentration of 100 μM using 6.0 eq. (ubiquitin) or 10.0 eq. of the LCT (hCA II) in 10 mM phosphate and 0.3 mM TCEP buffer with pH 7.0 at rt overnight. ¹H-¹⁵N HSQC and ¹H-¹⁵N HSQC IPAP spectra were measured in 10 mM phosphate buffer with pH 6.0 (ubiquitin S57C and ubiquitin K48C) and pH 6.8 (hCA II S166C) at a temperature of 298 K on a 600 MHz Bruker Avance III NMR spectrometer equipped with a cryogenic QCI-F probe. The obtained NMR spectra were assigned using CcpNmr Analysis.³ The tensor properties were then obtained by fitting the residues in secondary structure elements of ubiquitin (PDB 1UBI⁴ and 2MJB⁵) or the leucine residues of hCA II (PDB 3KS3⁶) using Numbat.⁷ The metal centres were found in a distance of 6.1-7.0 Å from the C_{beta} of the cysteine residue. PCS anisotropy parameters were derived exclusively from PCS data sets, while RDC anisotropy parameters were derived exclusively from RDC data sets with the metal position from the PCS data set as additional input. Q-factors were calculated using the following equation:

$$Q = \frac{\sqrt{\sum (PCS_{exp} - PCS_{calc})^2}}{\sqrt{\sum (PCS_{exp})^2}}$$

¹H and ¹³C NMR experiments were performed at a temperature of 298 K on a 600 MHz Bruker Avance III NMR spectrometer equipped with a cryogenic QCI-F probe. (2S,5S,8S,11S)-2,5,8,11-Tetraisopropyl-1,4,7,10-tetraazacyclododecane was synthesized according to the procedures described by Joss et al.⁸, 5-(bromomethyl)-2-(methylthio)thiazolo[5,4-*b*]pyridine according to the procedures published by Müntener et al.⁹ ESI-MS spectra were recorded on a Shimadzu LCMS-2020 liquid chromatograph mass spectrometer. HRMS spectra were measured on a Bruker MaXis 4G HiRes ESI Mass Spectrometer.

HPLC conditions

Analytical HPLC measurements were performed on a Shimadzu LC system (LC-20AT prominence liquid chromatograph, SIL-20A HT prominence auto sampler, CTO-20AC prominence column oven, CBM-20A communications bus module, SPD-20A prominence UV/VIS detector ($\lambda = 254$ and 280 nm, LC-20AD prominence liquid chromatograph) combined with a Shimadzu LCMS-2020 liquid chromatograph mass spectrometer. As column for analytical HPLC measurements, a ReproSil-Pur ODS, $3.3 \mu\text{m}$, 150×3 mm, provided by Maisch GmbH was used. Commercial HPLC grade solvents were used and a binary gradient was applied.

Solvent A: Milli-Q water + 0.1% TFA

Solvent B: 90% acetonitrile + 10% Milli-Q water + 0.085% TFA.

HPLC gradient: 95% A (min 0-2), linear gradient 95% A to 100% B (min 2-6), 100% B (min 6-14), linear gradient 100% B to 95% A (min 14-15), 95% A (min 15-22).

Preparative HPLC purification was performed on a Shimadzu LC system (LC-20AT prominence liquid chromatograph, SIL-20A HT prominence auto sampler, CTO-20AC prominence column oven, CBM-20A communications bus module, SPD-20A prominence UV/VIS detector ($\lambda = 254$ and 280 nm, LC-20AD prominence liquid chromatograph) combined with a Shimadzu LCMS-2020 liquid chromatograph mass spectrometer. As column for preparative HPLC purification, a ReproSil-Pur 120 ODS-3, $5 \mu\text{m}$, 150×20 mm, provided by Maisch GmbH was used. Commercial HPLC grade solvents were used and a binary gradient was applied during purification.

Solvent A: Milli-Q water + 0.1% TFA

Solvent B: 90% acetonitrile + 10% Milli-Q water + 0.085% TFA.

HPLC gradient: 95% A (min 0-2), linear gradient 95% A to 100% B (min 2-15), 100% B (min 15-22), linear gradient 100% B to 95% A (min 22-23), 95% A (min 23-25).

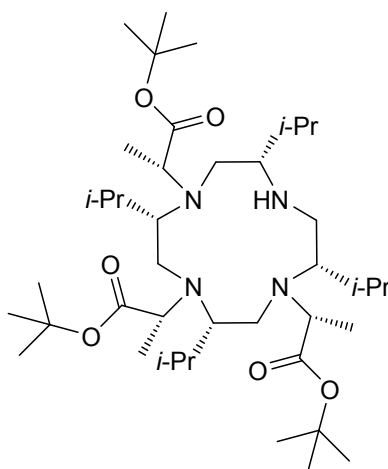
HPLC measurements of protein samples were performed using the direct injection mode on a Shimadzu LC system (LC-20AT prominence liquid chromatograph, SIL-20A HT prominence auto sampler, CTO-20AC prominence column oven, CBM-20A communications bus module, SPD-20A prominence UV/VIS detector ($\lambda = 254$ and 280 nm, LC-20AD prominence liquid chromatograph) combined with a Shimadzu LCMS-2020 liquid chromatograph mass spectrometer. Commercial HPLC grade solvents were used and a binary gradient was applied. MS spectra of proteins were deconvoluted using the Bruker Daltonics DataAnalysis software.

Solvent A: Milli-Q water + 0.1% TFA

Solvent B: 90% acetonitrile + 10% Milli-Q water + 0.085% TFA.

HPLC gradient: isocratic 95% A (min 0-4).

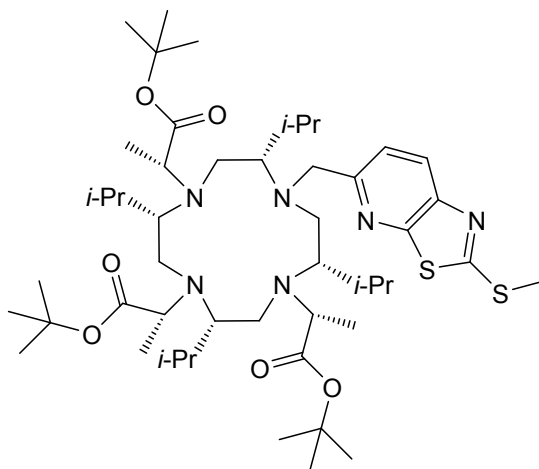
tri-*tert*-butyl 2,2',2''-((2*S*,5*S*,8*S*,11*S*)-2,5,8,11-tetraisopropyl-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)(2*R*,2'*R*,2''*R*)-tripropionate



tert-Butyl (*S*)-2-(((trifluoromethyl)sulfonyl)oxy)propanoate (245 mg, 0.882 mmol, 3.0 eq.) in acetonitrile (0.5 ml) was added to (2*S*,5*S*,8*S*,11*S*)-2,5,8,11-tetraisopropyl-1,4,7,10-tetraazacyclododecane (100 mg, 0.294 mmol, 1.0 eq.) and potassium carbonate (406 mg, 2.94 mmol, 10.0 eq.) in acetonitrile (10 ml) and the mixture was stirred at rt overnight. The reaction was quenched by addition of an excess triethylamine, the solvent was evaporated under reduced pressure and the residue subjected to preparative HPLC to give tri-*tert*-butyl 2,2',2''-((2*S*,5*S*,8*S*,11*S*)-2,5,8,11-tetraisopropyl-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)(2*R*,2'*R*,2''*R*)-tripropionate as white powder (133 mg, 62.4%).

HRMS: [M+H]⁺, C₄₁H₈₁N₄O₆, m/z (calc.) = 725.6151, m/z (meas.) = 725.6158.

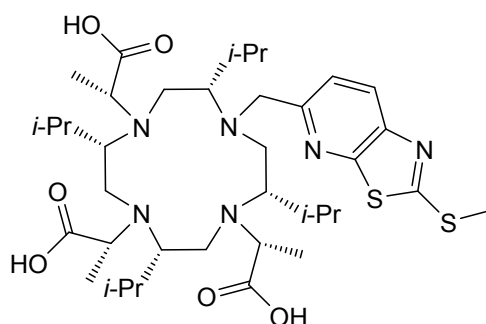
tri-*tert*-butyl 2,2',2''-((2*S*,5*S*,8*S*,11*S*)-2,5,8,11-tetraisopropyl-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)(2*R*,2'*R*,2''*R*)-tripropionate



5-(bromomethyl)-2-(methylthio)thiazolo[5,4-*b*]pyridine (125 mg, 0.455 mmol, 2.6 eq.) was added to tri-*tert*-butyl 2,2',2''-((2*S*,5*S*,8*S*,11*S*)-2,5,8,11-tetraisopropyl-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)(2*R*,2'*R*,2''*R*)-tripropionate (127 mg, 0.175, 1.0 eq.) in acetonitrile (10 ml) and the mixture was stirred for 4 h. Then potassium iodide (75.5 mg, 0.455 mmol, 2.6 eq.) was added and the mixture was stirred for 3 d at rt. The mixture was filtered and the solvent evaporated under reduced pressure. The residue was subjected to preparative HPLC to give tri-*tert*-butyl 2,2',2''-((2*S*,5*S*,8*S*,11*S*)-2,5,8,11-tetraisopropyl-10-((2-(methylthio)thiazolo[5,4-*b*]pyridin-5-yl)methyl)-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)(2*R*,2'*R*,2''*R*)-tripropionate as white powder (75 mg, 46.6%).

HRMS: [M+H]⁺, C₄₉H₈₇N₆O₆S₂, m/z (calc.) = 919.6123, m/z (meas.) = 919.6115.

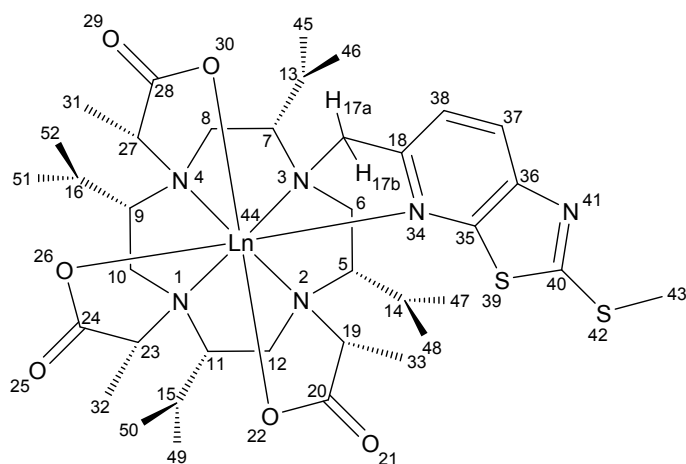
(2*R*,2'*R*,2''*R*)-2,2',2''-((2*S*,5*S*,8*S*,11*S*)-2,5,8,11-tetraisopropyl-10-((2-(methylthio)thiazolo[5,4-*b*]pyridin-5-yl)methyl)-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)tripropionic acid



tri-*tert*-butyl 2,2',2''-((2*S*,5*S*,8*S*,11*S*)-2,5,8,11-tetraisopropyl-10-((2-(methylthio)thiazolo[5,4-*b*]pyridin-5-yl)methyl)-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)(2*R*,2'*R*,2''*R*)-tripropionate (64.0 mg, 0.070 mmol, 1.0 eq.) was dissolved in a mixture of TFA, thioanisole and water (92:6:2, 5 ml) and stirred for 5 h at rt. Then the mixture was heated to 50 °C for 30 min and solvents were evaporated under reduced pressure. The residue was subjected to preparative HPLC to give (2*R*,2'*R*,2''*R*)-2,2',2''-((2*S*,5*S*,8*S*,11*S*)-2,5,8,11-tetraisopropyl-10-((2-(methylthio)thiazolo[5,4-*b*]pyridin-5-yl)methyl)-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)tripropionic acid as a colourless glassy solid (40 mg, 76.3%).

HRMS: [M+H]⁺, C₃₇H₆₃N₆O₆S₂, m/z (calc.) = 751.4245, m/z (meas.) = 751.4246.

Ln-(2*R*,2'*R*,2''*R*)-2,2',2''-((2*S*,5*S*,8*S*,11*S*)-2,5,8,11-tetraisopropyl-10-((2-(methylthio)thiazolo[5,4-*b*]pyridin-5-yl)methyl)-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)tripropionate, Ln-P4T-DOTA_{red}



Lanthanide trifluoromethanesulfonate (2.0 eq.) was added to (2*R*,2'*R*,2''*R*)-2,2',2''-((2*S*,5*S*,8*S*,11*S*)-2,5,8,11-tetraisopropyl-10-((2-(methylthio)thiazolo[5,4-*b*]pyridin-5-yl)methyl)-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)tripropionic acid (1.0 eq.) in aq. ammonium acetate (10 ml, 100 mM) and stirred overnight at 80 °C. The solvent was evaporated and the residue was subjected to preparative HPLC to give Ln-(2*R*,2'*R*,2''*R*)-2,2',2''-((2*S*,5*S*,8*S*,11*S*)-2,5,8,11-tetraisopropyl-10-((2-(methylthio)thiazolo[5,4-*b*]pyridin-5-yl)methyl)-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)tripropionate as white powders (Lutetium: 12 mg, 75.1%, Thulium: 10 mg, 63.0%, Dysprosium: 7 mg, 44.4%).

$\Lambda(\delta\delta\delta\delta)$ -Lu-P4T-DOTA_{red}

¹H NMR (600 MHz, Deuterium Oxide, pH 6.0) δ 8.25 (d, *J* = 8.5 Hz, 1H, H37), 7.54 (d, *J* = 8.5 Hz, 1H, H38), 4.70 (d, *J* = 17.5 Hz, 1H, H17b), 4.43 (d, *J* = 17.4 Hz, 1H, H17a), 3.84 (q, *J* = 7.4 Hz, 1H, H27), 3.76 (q, *J* = 7.3 Hz, 1H, H23), 3.47 (dd, *J* = 13.7, 13.7 Hz, H6ax), 3.18 - 3.06 (m, 2H, H8eq,8ax), 3.06 - 2.98 (m, 3H, H12ax,10eq,10ax), 2.98 - 2.92 (m, 3H, H7,12eq,6eq), 2.88 - 2.84 (m, 1H, H9), 2.81 (q, *J* = 7.0 Hz, 3H, H19), 2.78 (s, 3H, H43), 2.71 - 2.65 (m, 2H, H5,11), 2.54 - 2.47 (m, 2H, H15,16), 2.47 - 2.41 (m, 1H, H13), 2.37 - 2.30 (m, 1H, H14), 1.59 (d, *J* = 7.2 Hz, 3H, H32), 1.58 (d, *J* = 7.2 Hz, 3H, H31), 1.30 (d, *J* = 7.0 Hz, 3H, H33), 1.12 (d, *J* = 7.1 Hz, 3H, H46), 1.10 (d, *J* = 6.8 Hz, 3H, H45), 1.07 (d, *J* = 7.1 Hz, 3H, H52), 1.04 (d, *J* = 7.1 Hz, 3H, H51), 1.03 (d, *J* = 7.5 Hz, 3H, H50), 1.01 (d, *J* = 6.8 Hz, 3H, H49), 0.95 (d, *J* = 6.8 Hz, 3H, H47), 0.84 (d, *J* = 6.9 Hz, 3H, H48).

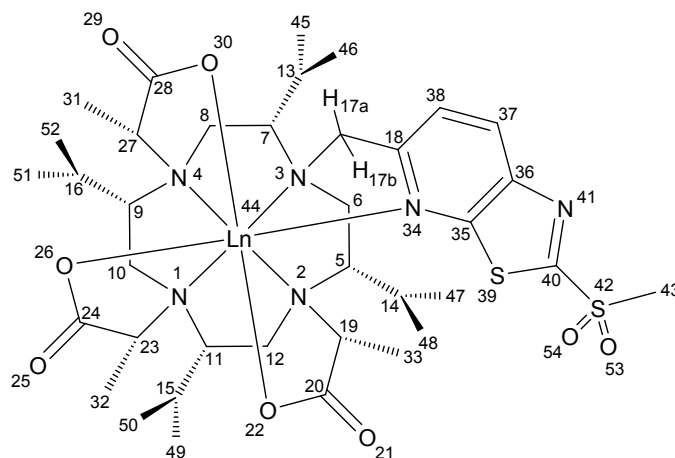
¹³C NMR (151 MHz, Deuterium Oxide, pH 6.0, TSP-*d*₄, extracted from HMQC and HMBC) δ 186.1 (C28), 185.8 (C24), 185.4 (C20), 177.6 (C40), 158.7 (C35), 158.7 (C18), 150.5 (C36), 134.4 (C37), 123.0 (C38), 73.4 (C7), 73.1 (C11), 73.1 (C9), 72.8 (C5), 71.1 (C23), 70.9 (C27), 69.6 (C19), 67.7 (C17), 53.1 (C6), 45.1 (C10), 44.8 (C8), 44.4 (C12), 29.9 (C16), 29.9 (C15), 29.4 (C14), 27.4 (C13), 26.7 (C50), 26.7 (C46), 26.5 (C52), 25.3 (C48), 20.4 (C45), 20.0 (C49), 20.0 (C47), 20.0 (C51), 18.2 (C43), 17.0 (C31), 16.7 (C32), 16.2 (C33).

Ln = Lu, HRMS: [M+H]⁺, C₃₇H₆₀LuN₆O₆S₂, *m/z* (calc.) = 923.3418, *m/z* (meas.) = 923.3418.

Ln = Tm, HRMS: [M+H]⁺, C₃₇H₆₀N₆O₆S₂Tm, *m/z* (calc.) = 917.3352, *m/z* (meas.) = 917.3356.

Ln = Dy, HRMS: [M+H]⁺, C₃₇H₆₀DyLuN₆O₆S₂, *m/z* (calc.) = 912.3302, *m/z* (meas.) = 912.3302.

$\Lambda(\delta\delta\delta\delta)$ -Ln-(2*R*,2'*R*,2''*R*)-2,2',2''-((2*S*,5*S*,8*S*,11*S*)-2,5,8,11-tetraisopropyl-10-((2-(methylsulfonyl)thiazolo[5,4-*b*]pyridin-5-yl)methyl)-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)tripropionate, Ln-P4T-DOTA



meta-Chloroperbenzoic acid (10.0 eq.) was added to Ln-(2*R*,2'*R*,2''*R*)-2,2',2''-((2*S*,5*S*,8*S*,11*S*)-2,5,8,11-tetraisopropyl-10-((2-(methylthio)thiazolo[5,4-*b*]pyridin-5-yl)methyl)-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)tripropionate (1.0 eq.) in DCM (10 ml) and the mixture was stirred overnight at rt. Then *meta*-chloroperbenzoic acid (10.0 eq.) was added and the mixture was stirred for 3 h at rt. The mixture was washed with sodium bicarbonate and the organic phase was subjected to preparative HPLC to give $\Lambda(\delta\delta\delta\delta)$ -Ln-(2*R*,2'*R*,2''*R*)-2,2',2''-((2*S*,5*S*,8*S*,11*S*)-2,5,8,11-tetraisopropyl-10-((2-(methylsulfonyl)thiazolo[5,4-*b*]pyridin-5-yl)methyl)-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)tripropionate as white powders (Lutetium: 3.5 mg, 28.2%, Thulium: 3.2 mg, 28.0%, Dysprosium: 2.5 mg, 34.5%). Solutions of Ln-P4T-DOTA were stored in dry acetonitrile after purification by preparative HPLC.

$\Lambda(\delta\delta\delta\delta)$ -Lu-P4T-DOTA

$^1\text{H NMR}$ (600 MHz, Deuterium Oxide, pH 6.0) δ 8.73 (d, *J* = 8.7 Hz, 1H, H37), 7.81 (d, *J* = 8.8 Hz, 1H, H38), 4.87 (d, *J* = 18.7 Hz, 1H, H17b), 4.60 (d, *J* = 18.5 Hz, 1H, H17a), 3.85 (q, *J* = 7.5 Hz, 1H, H27), 3.78 (q, *J* = 7.7 Hz, 1H, H23), 3.56 - 3.48 (m, H43,6ax), 3.18 - 3.07 (m, 3H, H8ax,10eq,12ax), 3.07 - 3.01 (m, 2H, H10ax,12eq), 3.01 - 2.91 (m, 4H, H7,8eq,6eq,19), 2.90 - 2.84 (m, 1H, H9), 2.71 - 2.68 (m, 1H, H11), 2.67 - 2.64 (m, 1H, H5), 2.55 - 2.41 (m, 3H, H15,16,13), 2.41 - 2.30 (m, 1H, H14), 1.60 (d, *J* = 7.4 Hz, 3H, H32), 1.57 (d, *J* = 7.3 Hz, 3H, H31), 1.32 (d, *J* = 7.1 Hz, 3H, H33), 1.13 (d, *J* = 7.4 Hz, 3H, H46), 1.11 (d, *J* = 6.9 Hz, 3H, H45), 1.07 (d, *J* = 7.1 Hz, 3H, H52), 1.04 (d, *J* = 6.4 Hz, 3H, H51), 1.03 (d, *J* = 6.7 Hz, 3H, H50), 1.01 (d, *J* = 6.8 Hz, 3H, H49), 0.96 (d, *J* = 6.7 Hz, 3H, H47), 0.85 (d, *J* = 6.9 Hz, 3H, H48).

$^{13}\text{C NMR}$ (151 MHz, Deuterium Oxide, pH 6.0, TSP- d_4 , extracted from HMQC and HMBC) δ 186.3 (C28), 185.7 (C24), 185.4 (C20), 174.7 (C40), 164.7 (C18), 158.9 (C35), 149.7 (C36), 140.1 (C37), 125.1 (C38), 73.6 (C7), 73.4 (C11), 73.3 (C9), 73.2 (C5), 71.3 (C23), 71.0 (C27), 69.8 (C19), 68.3 (C17), 53.4 (C6), 45.2 (C12), 45.0 (C43), 45.0 (C10), 44.6 (C8), 30.0 (C16), 30.0 (C15), 29.5 (C14), 27.7 (C13), 26.7 (C46), 26.6 (C52), 25.4 (C48), 20.4 (C45), 20.1 (C47), 20.1 (C50), 20.1 (C49), 17.1 (C32), 17.1 (C31), 16.4 (C33).

Ln = Lu, HRMS: $[\text{M}+\text{H}]^+$, $\text{C}_{37}\text{H}_{60}\text{LuN}_6\text{O}_8\text{S}_2$, m/z (calc.) = 955.3316, m/z (meas.) = 955.3322.

Ln = Tm, HRMS: $[\text{M}+\text{H}]^+$, $\text{C}_{37}\text{H}_{60}\text{N}_6\text{O}_8\text{S}_2\text{Tm}$, m/z (calc.) = 949.3251, m/z (meas.) = 949.3266.

Ln = Dy, HRMS: $[\text{M}+\text{H}]^+$, $\text{C}_{37}\text{H}_{60}\text{DyLuN}_6\text{O}_8\text{S}_2$, m/z (calc.) = 944.3200, m/z (meas.) = 944.3210.

¹H-NMR of Lu-P4T-DOTA_{red}

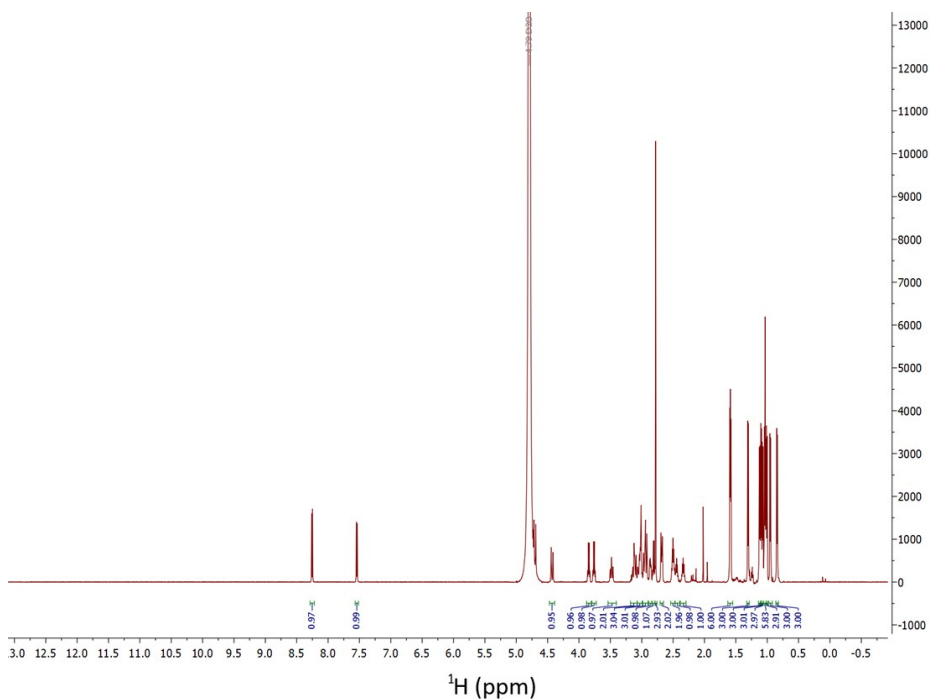


Figure S1: ¹H-NMR of Lu-P4T-DOTA_{red} in D₂O (pH 6.0) at 298 K.

¹H-NMR of Lu-P4T-DOTA

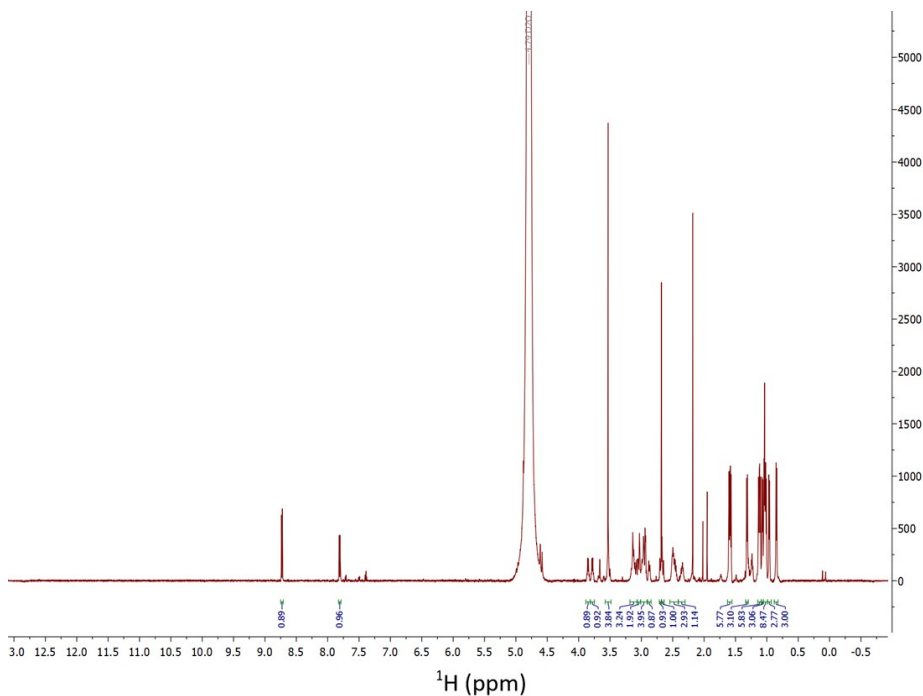


Figure S2: ¹H-NMR of Lu-P4T-DOTA in D₂O (pH 6.0) at 298 K.

¹H-NMR of Tm-P4T-DOTA

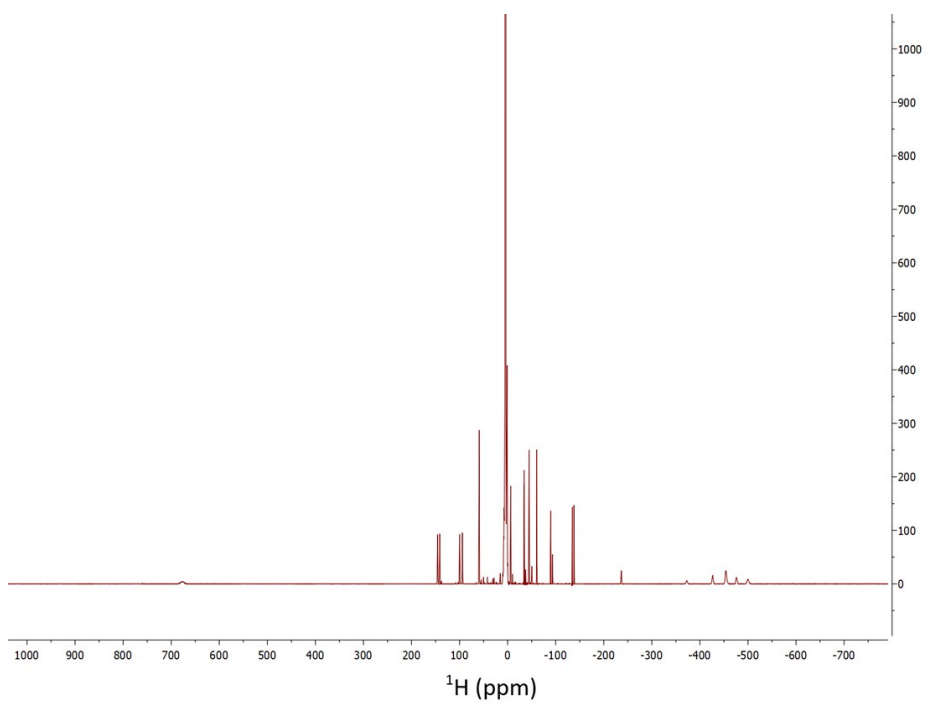


Figure S3: ¹H-NMR of Tm-P4T in D₂O at 298 K.

¹H-NMR of Dy-P4T-DOTA

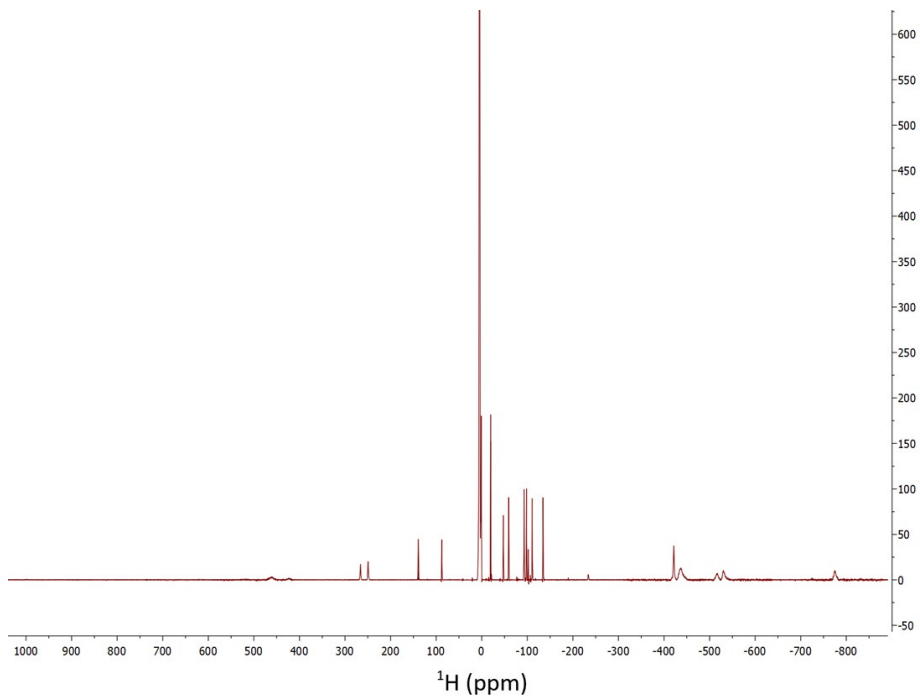


Figure S4: ¹H-NMR of Dy-P4T in D₂O at 298 K.

HPLC measurements of Ln-P4T

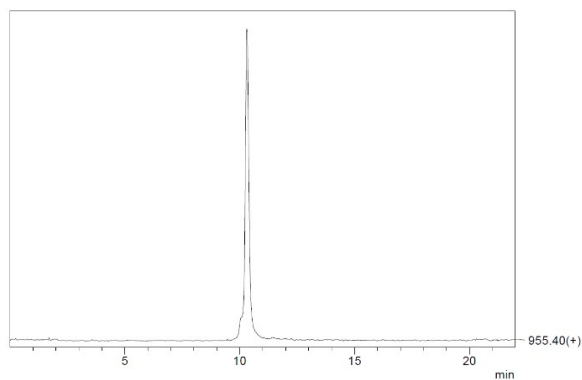


Figure S5: HPLC-MS trace of Lu-P4T.

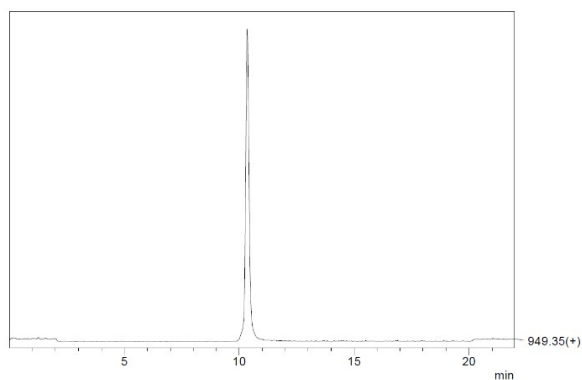


Figure S6: HPLC-MS trace of Tm-P4T.

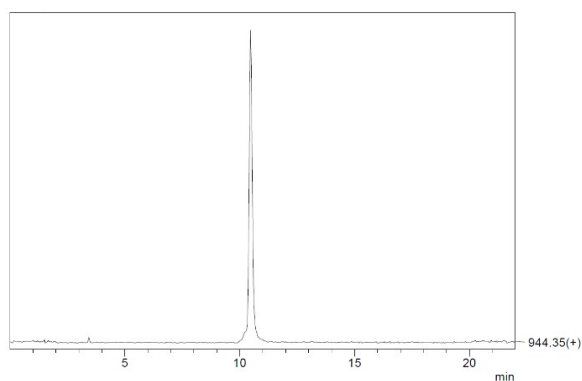


Figure S7: HPLC-MS trace of Dy-P4T.

Overlay of ^1H - ^{15}N HSQC spectra of Dy-P4T-Ub^{S57C} (blue), Tm-P4T-Ub^{S57C} (red) and Lu-P4T-Ub^{S57C} (black)

Measured in 10 mM phosphate buffer with pH 6.0 at a temperature of 298 K on a 600 MHz Bruker Avance III NMR spectrometer equipped with a cryogenic QCI-F probe.

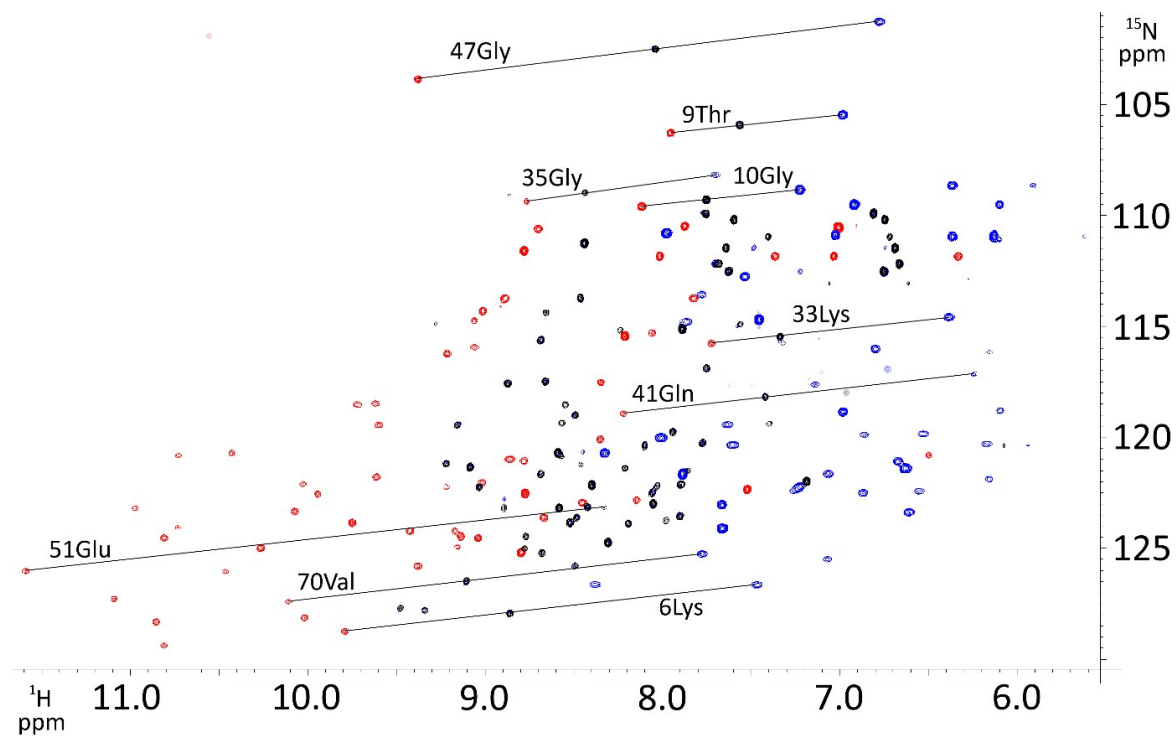


Figure S8: Overlay of ^1H - ^{15}N HSQC spectra of Dy-P4T-Ub^{S57C} (blue), Tm-P4T-Ub^{S57C} (red) and Lu-P4T-Ub^{S57C} (black).

Overlay of ^1H - ^{15}N HSQC spectra of Dy-P4T-Ub^{K48C} (blue), Tm-P4T-Ub^{K48C} (red) and Lu-P4T-Ub^{K48C} (black) attached to ubiquitin K48C

Measured in 10 mM phosphate buffer with pH 6.0 at a temperature of 298 K on a 600 MHz Bruker Avance III NMR spectrometer equipped with a cryogenic QCI-F probe.

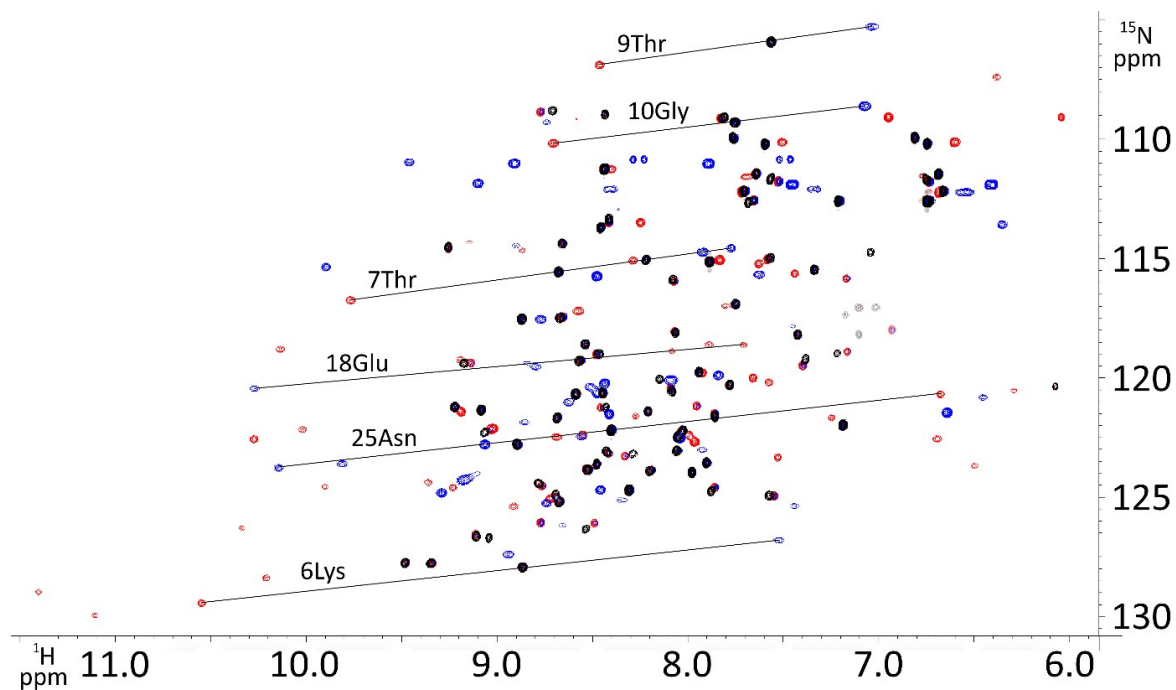


Figure S9: Overlay of ^1H - ^{15}N HSQC spectra of Dy-P4T-Ub^{K48C} (blue), Tm-P4T-Ub^{K48C} (red) and Lu-P4T-Ub^{K48C} (black).

Overlay of ^1H - ^{15}N HSQC spectra of Dy-P4T (blue), Tm-P4T (red) and Lu-P4T (black) attached to selectively ^{15}N leucine labeled hCA II S166C

Measured in 10 mM phosphate buffer with pH 6.8 at a temperature of 298 K on a 600 MHz Bruker Avance III NMR spectrometer equipped with a cryogenic QCI-F probe.

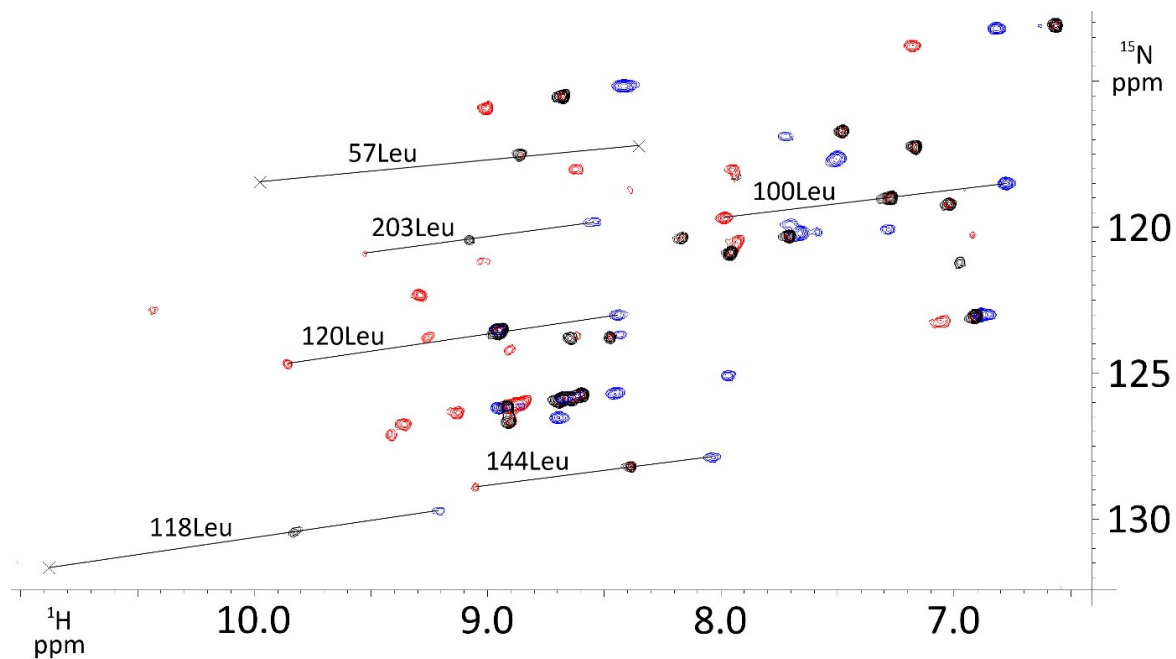


Figure S10: Overlay of ^1H - ^{15}N HSQC spectra of Dy-P4T (blue), Tm-P4T (red) and Lu-P4T (black) attached to selectively ^{15}N leucine labeled hCA II S166C.

Shift list comparison of ^1H - ^{15}N HSQC spectra of Dy- and Lu-P4T-Ub^{S57C}

Table S1: Shift list comparison of ^1H - ^{15}N HSQC spectra of Dy- and Lu-P4T-Ub^{S57C}.

Residue	Reson. 1	Shift Dy	Shift Lu	Reson. 2		Shift Dy	Shift Lu
2Gln	H	6.18	8.89	N		120.31	123.17
4Phe	H	6.16	8.55	N		116.17	118.53
5Val	H	7.63	9.22	N		119.43	121.19
6Lys	H	7.47	8.86	N		126.64	127.93
12Thr	H	8.01	8.59	N		120.02	120.71
13Ile	H	8.38	9.48	N		126.63	127.69
14Thr	H	7.61	8.69	N		120.35	121.66
15Leu	H	6.61	8.68	N		123.37	125.22
28Ala	H	6.16	7.90	N		121.88	123.56
29Lys	H	6.10	7.77	N		118.80	120.26
30Ile	H	6.53	8.21	N		119.84	121.39
31Gln	H	7.23	8.48	N		122.27	123.61
32Asp	H	6.98	7.94	N		118.85	119.76
33Lys	H	6.39	7.34	N		114.59	115.46
34Glu	H	7.78	8.66	N		113.56	114.38
39Asp	H	7.54	8.46	N		112.76	113.72
40Gln	H	6.80	7.75	N		116.01	116.90
41Gln	H	6.24	7.42	N		117.14	118.18
42Arg	H	7.06	8.42	N		121.65	123.15
43Leu	H	6.55	8.77	N		122.43	124.47
44Ile	H	6.86	9.03	N		119.88	122.25
49Gln	H	6.67	8.58	N		121.08	123.19
67Leu	H	7.07	9.34	N		125.49	127.80
69Leu	H	6.87	8.19	N		122.50	123.88
70Val	H	7.78	9.11	N		125.26	126.48

Shift list comparison of ^1H - ^{15}N HSQC spectra of Tm- and Lu-P4T-Ub^{S57C}

Table S2: Shift list comparison of ^1H - ^{15}N HSQC spectra of Tm- and Lu-P4T-Ub^{S57C}.

Residue	Reson. 1	Shift Tm	Shift Lu	Reson. 2	Shift Tm	Shift Lu
3Ile	H	9.06	8.24	N	115.95	115.17
4Phe	H	9.60	8.55	N	119.45	118.53
5Val	H	10.03	9.22	N	122.11	121.19
6Lys	H	9.79	8.86	N	128.74	127.93
7Thr	H	9.21	8.69	N	116.23	115.63
12Thr	H	8.86	8.59	N	120.99	120.71
13Ile	H	10.02	9.48	N	128.14	127.69
14Thr	H	9.02	8.69	N	122.07	121.66
15Leu	H	9.38	8.68	N	125.80	125.22
16Glu	H	8.45	8.06	N	122.94	122.50
23Ile	H	13.34	8.46	N	125.72	121.22
25Asn	H	10.07	7.86	N	123.34	121.51
26Val	H	10.81	8.03	N	124.53	122.21
27Lys	H	10.43	8.49	N	120.72	119.00
28Ala	H	9.04	7.90	N	124.54	123.56
29Lys	H	8.78	7.77	N	121.06	120.26
30Ile	H	9.22	8.21	N	122.24	121.39
31Gln	H	9.17	8.48	N	124.22	123.61
32Asp	H	8.35	7.94	N	120.09	119.76
33Lys	H	7.72	7.34	N	115.76	115.46
34Glu	H	9.06	8.66	N	114.75	114.38
39Asp	H	9.01	8.46	N	114.31	113.72
40Gln	H	8.35	7.75	N	117.53	116.90
41Gln	H	8.22	7.42	N	118.93	118.18
42Arg	H	9.42	8.42	N	124.22	123.15
43Leu	H	10.46	8.77	N	126.05	124.47
44Ile	H	10.73	9.03	N	124.06	122.25
45Phe	H	11.09	8.78	N	127.28	125.01
48Lys	H	9.75	7.90	N	123.86	122.12
49Gln	H	10.26	8.58	N	124.99	123.19
66Thr	H	9.72	8.66	N	118.54	117.48
67Leu	H	10.81	9.34	N	129.38	127.80
68His	H	10.73	9.16	N	120.82	119.45
69Leu	H	9.15	8.66	N	124.93	123.87
70Val	H	10.11	9.11	N	127.40	126.48
71Leu	H	8.67	8.05	N	123.62	123.00

Shift list comparison of ^1H - ^{15}N HSQC spectra of Dy- and Lu-P4T-Ub^{K48C}

Table S3: Shift list comparison of ^1H - ^{15}N HSQC spectra of Dy- and Lu-P4T-Ub^{K48C}.

Residue	Reson. 1	Shift Dy	Shift Lu	Reson. 2	Shift Dy	Shift Lu
2Gln	H	8.56	8.89	N	122.46	122.78
3Ile	H	7.92	8.22	N	114.73	115.07
5Val	H	8.44	9.22	N	120.22	121.22
6Lys	H	7.52	8.87	N	126.81	127.94
7Thr	H	7.77	8.68	N	114.57	115.56
13Ile	H	8.94	9.48	N	127.41	127.76
14Thr	H	8.45	8.68	N	121.25	121.68
27Lys	H	8.80	8.46	N	119.53	119.00
30Ile	H	8.41	8.21	N	121.53	121.41
32Asp	H	8.09	7.94	N	120.10	119.76
33Lys	H	7.63	7.34	N	115.67	115.47
39Asp	H	8.90	8.46	N	114.46	113.71
40Gln	H	8.77	7.75	N	117.55	116.91
42Arg	H	9.17	8.43	N	124.29	123.08
70Val	H	8.66	9.11	N	126.18	126.65

Shift list comparison of ^1H - ^{15}N HSQC spectra of Tm- and Lu-P4T-Ub^{K48C}

Table S4: Shift list comparison of ^1H - ^{15}N HSQC spectra of Tm- and Lu-P4T-Ub^{K48C}.

Residue	Reson. 1	Shift Tm	Shift Lu	Reson. 2	Shift Tm	Shift Lu
2Gln	H	8.69	8.89	N	122.48	122.78
3Ile	H	8.29	8.22	N	115.08	115.07
5Val	H	10.02	9.22	N	122.17	121.22
6Lys	H	10.55	8.87	N	129.44	127.94
7Thr	H	9.77	8.68	N	116.75	115.56
12Thr	H	9.19	8.59	N	121.43	120.68
13Ile	H	10.21	9.48	N	128.39	127.76
25Asn	H	6.68	7.86	N	120.69	121.62
26Val	H	7.25	8.03	N	121.68	122.23
27Lys	H	8.08	8.46	N	118.88	119.00
28Ala	H	7.53	7.90	N	123.33	123.56
29Lys	H	7.57	7.78	N	120.19	120.30
30Ile	H	8.27	8.21	N	121.60	121.41
33Lys	H	7.44	7.34	N	115.64	115.47
34Glu	H	8.87	8.66	N	114.67	114.38
39Asp	H	8.25	8.44	N	113.49	113.53
40Gln	H	7.81	7.75	N	116.98	116.91
41Gln	H	7.89	7.42	N	118.61	118.19
42Arg	H	9.90	8.43	N	124.56	123.08
43Leu	H	11.57	8.78	N	127.37	124.43
44Ile	H	13.95	9.06	N	128.26	122.29
66Thr	H	10.14	8.67	N	118.80	117.48
67Leu	H	11.11	9.34	N	129.95	127.77
68His	H	13.29	9.17	N	122.85	119.40
69Leu	H	10.34	8.20	N	126.31	123.90
70Val	H	11.40	9.11	N	128.98	126.65
71Leu	H	9.36	8.06	N	124.38	123.06

Shift list comparison of ^1H - ^{15}N HSQC spectra of Dy- and Lu-P4T attached to selectively ^{15}N leucine labelled hCA II S166C

Table S5: Shift list comparison of ^1H - ^{15}N HSQC spectra of Dy- and Lu-P4T attached to selectively ^{15}N leucine labelled hCA II S166C.

Residue	Reson. 1	Shift Dy	Shift Lu	Reson. 2	Shift Dy	Shift Lu
47Leu	H	8.41	8.68	N	115.17	115.53
57Leu	H	6.82	6.56	N	113.19	113.10
79Leu	H	7.72	7.48	N	116.90	116.72
84Leu	H	7.51	7.17	N	117.67	117.26
90Leu	H	6.78	7.27	N	118.51	119.01
100Leu	H	8.35	8.86	N	117.23	117.52
118Leu	H	8.54	9.08	N	119.83	120.44
120Leu	H	8.97	8.95	N	123.50	123.55
141Leu	H	8.44	8.95	N	123.02	123.56
144Leu	H	8.43	8.47	N	123.69	123.77
148Leu	H	9.21	9.83	N	129.72	130.44
157Leu	H	8.03	8.39	N	127.89	128.21
184Leu	H	7.49	7.93	N	117.67	118.30
185Leu	H	7.71	8.16	N	119.90	120.36
198Leu	H	8.70	8.91	N	126.53	126.65
203Leu	H	7.97	8.59	N	125.09	125.74
212Leu	H	8.45	8.64	N	125.71	125.86
224Leu	H	7.28	7.96	N	120.09	120.88
240Leu	H	7.66	7.71	N	120.21	120.32
251Leu	H	8.95	8.91	N	126.20	126.17

Shift list comparison of ^1H - ^{15}N HSQC spectra of Tm- and Lu-P4T attached to selectively ^{15}N leucine labelled hCA II S166C

Table S6: Shift list comparison of ^1H - ^{15}N HSQC spectra of Tm- and Lu-P4T attached to selectively ^{15}N leucine labelled hCA II S166C.

Residue	Reson. 1	Shift Tm	Shift Lu	Reson. 2	Shift Tm	Shift Lu
44Leu	H	7.06	6.91	N	123.24	123.06
47Leu	H	9.13	8.91	N	126.37	126.17
57Leu	H	9.98	8.86	N	118.47	117.52
79Leu	H	9.26	8.95	N	123.81	123.55
84Leu	H	7.93	7.71	N	120.54	120.32
90Leu	H	8.91	8.47	N	124.25	123.77
100Leu	H	7.98	7.27	N	119.69	119.01
118Leu	H	10.88	9.83	N	131.67	130.44
120Leu	H	9.86	8.95	N	124.69	123.56
141Leu	H	9.00	8.68	N	115.94	115.53
144Leu	H	9.05	8.39	N	128.90	128.21
148Leu	H	9.03	8.16	N	121.17	120.36
157Leu	H	8.63	7.48	N	118.02	116.72
184Leu	H	7.95	7.17	N	118.04	117.26
185Leu	H	7.18	6.56	N	113.79	113.10
189Leu	H	8.87	8.69	N	126.06	125.92
198Leu	H	8.39	7.93	N	118.72	118.30
203Leu	H	9.53	9.08	N	120.91	120.44
212Leu	H	9.42	8.91	N	127.12	126.65
224Leu	H	9.29	7.96	N	122.33	120.88
229Leu	H	10.44	7.02	N	122.85	119.20
240Leu	H	9.36	8.59	N	126.75	125.74
251Leu	H	8.84	8.64	N	125.95	125.86

Residual dipolar couplings measured in ^1H - ^{15}N HSQC spectra of Dy- and Lu-P4T-Ub^{S57C}

Table S7: Residual dipolar couplings measured in ^1H - ^{15}N HSQC spectra of Dy- and Lu-P4T-Ub^{S57C}.

Residue	RDC (Hz)
2	9.5
4	1.8
5	1.2
6	5.1
12	-1.1
13	-0.6
14	5.5
15	-3.9
28	11.1
29	11.4
30	5.8
32	14.3
33	8.7
34	5.7
39	-8.4
40	-4.5
41	-28.5
42	-8.3
43	-8.6
44	1.3
49	-0.5
67	-3.8
69	-5.1
70	-11.1

Residual dipolar couplings measured in ^1H - ^{15}N HSQC spectra of Tm- and Lu-P4T-Ub^{S57C}

Table S8: Residual dipolar couplings measured in ^1H - ^{15}N HSQC spectra of Tm- and Lu-P4T-Ub^{S57C}.

Residue	RDC (Hz)
3	12.7
4	9.9
5	6.3
6	5.5
7	2.4
12	3.6
13	3.2
14	8.0
15	11.6
16	7.9
23	8.6
25	1.6
26	5.5
27	5.9
28	-6.9
29	9.4
30	6.0
31	1.4
32	-0.5
33	5.0
34	3.6
39	14.1
40	-4.8
41	5.6
42	3.0
43	7.1
44	4.1
45	2.8
48	-14.7
49	-2.9
66	11.9
67	4.5
68	5.9
69	8.2
70	2.0
71	2.6

Correlation plots of experimental and back-calculated PCS and RDC

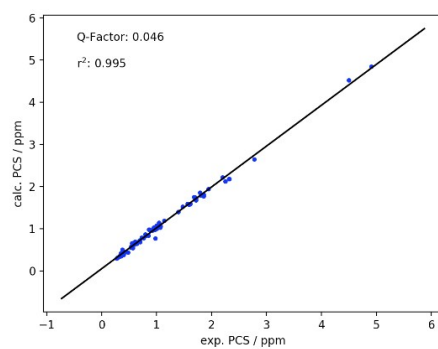


Figure S11: PCS correlation plot of ^{15}N labelled ubiquitin S57C labelled with Tm-P4T-DOTA.

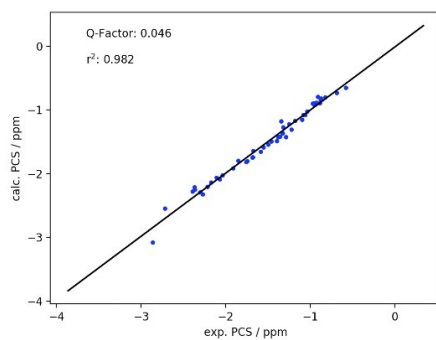


Figure S12: PCS correlation plot of ^{15}N labelled ubiquitin S57C labelled with Dy-P4T-DOTA.

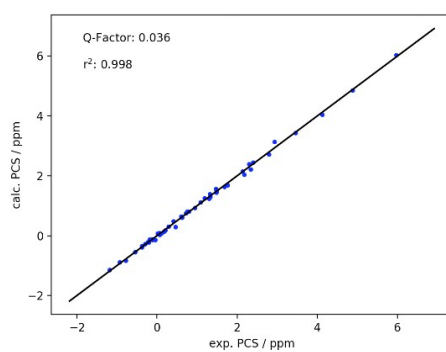


Figure S13: PCS correlation plot of ^{15}N labelled ubiquitin K48C labelled with Tm-P4T-DOTA.

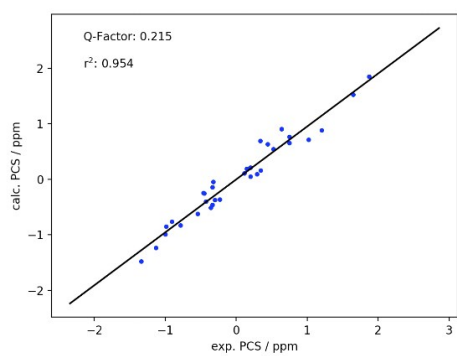


Figure S14: PCS correlation plot of ^{15}N labelled ubiquitin K48C labelled with Dy-P4T-DOTA.

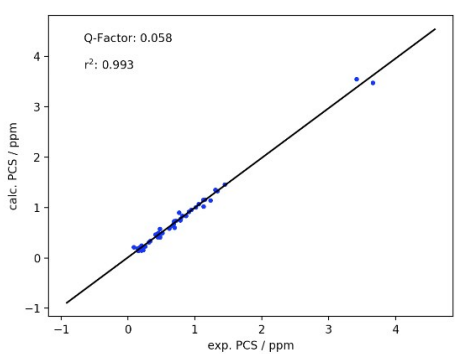


Figure S15: PCS correlation plot of selectively ^{15}N leucine labelled hCA S166C labelled with Tm-P4T-DOTA.

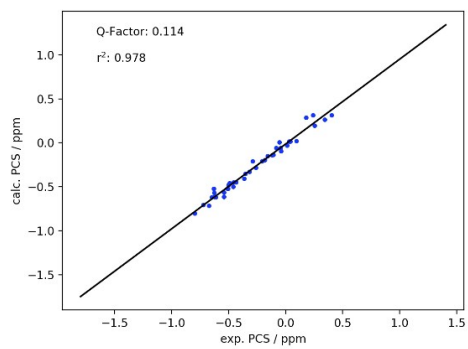


Figure S16: PCS correlation plot of selectively ^{15}N leucine labelled hCA S166C labelled with Dy-P4T-DOTA.

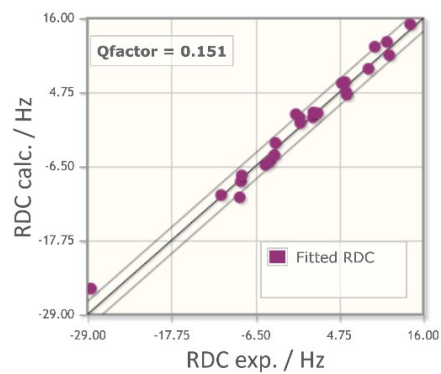


Figure S17: RDC correlation plot of ^{15}N labelled ubiquitin S57C labelled with Tm-P4T-DOTA.

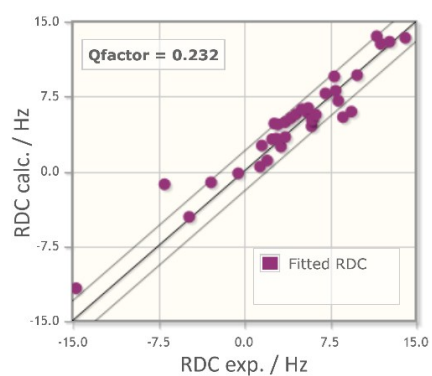


Figure S18: RDC correlation plot of ^{15}N labelled ubiquitin S57C labelled with Dy-P4T-DOTA.

DFT calculations of the Ln-P4T complexes in vacuo and water solvent

For the calculations, BP86 was used as functional,¹⁰⁻¹¹ SARC-TZVP as basis set for the ligands,¹² while SARC2-QZVP was used as basis set for the lanthanide metal.¹³ The calculations were performed using the relativistic ZORA approximation,¹⁴ as well as the RI approximation to speed up the calculations.¹⁵⁻¹⁶ To model the water solvent, CPCM solvent model was implemented into the calculations.¹⁷

References

1. Sass, J.; Cordier, F.; Hoffmann, A.; Rogowski, M.; Cousin, A.; Omichinski, J. G.; Löwen, H.; Grzesiek, S., Purple Membrane Induced Alignment of Biological Macromolecules in the Magnetic Field. *JACS* **1999**, *121* (10), 2047-2055.
2. Varghese, S.; Halling, P. J.; Häussinger, D.; Wimperis, S., High-Resolution Structural Characterization of a Heterogeneous Biocatalyst Using Solid-State NMR. *J. Phys. Chem. C* **2016**, *120* (50), 28717-28726.
3. Vranken, W. F.; Boucher, W.; Stevens, T. J.; Fogh, R. H.; Pajon, A.; Llinas, M.; Ulrich, E. L.; Markley, J. L.; Ionides, J.; Laue, E. D., The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins: Struct., Funct., Bioinf.* **2005**, *59* (4), 687-696.
4. Ramage, R.; Green, J.; Muir, T. W.; Ogunjobi, O. M.; Love, S.; Shaw, K., Synthetic, structural and biological studies of the ubiquitin system: the total chemical synthesis of ubiquitin. *Biochem. J* **1994**, *299* (Pt 1), 151-8.
5. Maltsev, A. S.; Grishaev, A.; Roche, J.; Zasloff, M.; Bax, A., Improved Cross Validation of a Static Ubiquitin Structure Derived from High Precision Residual Dipolar Couplings Measured in a Drug-Based Liquid Crystalline Phase. *JACS* **2014**, *136* (10), 3752-3755.
6. Avvaru, B. S.; Kim, C. U.; Sippel, K. H.; Gruner, S. M.; Agbandje-McKenna, M.; Silverman, D. N.; McKenna, R., A short, strong hydrogen bond in the active site of human carbonic anhydrase II. *Biochemistry* **2010**, *49* (2), 249-51.
7. Schmitz, C.; Stanton-Cook, M. J.; Su, X.-C.; Otting, G.; Huber, T., Numbat: an interactive software tool for fitting $\Delta\chi$ -tensors to molecular coordinates using pseudocontact shifts. *J. Biomol. NMR* **2008**, *41* (3), 179.
8. Joss, D.; Bertrams, M.-S.; Häussinger, D., A sterically overcrowded, isopropyl-substituted lanthanide chelating tag for protein PCS NMR spectroscopy: Synthesis of its macrocyclic scaffold and benchmarking on ubiquitin S57C and hCA II S166C. *Chem. Eur. J.* **2019**, *in press*, DOI:10.1002/chem.201901692.
9. Müntener, T.; Kottelat, J.; Huber, A.; Häussinger, D., New Lanthanide Chelating Tags for PCS NMR Spectroscopy with Reduction Stable, Rigid Linkers for Fast and Irreversible Conjugation to Proteins. *Bioconjugate Chem.* **2018**, *29* (10), 3344-3351.
10. Becke, A. D., Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988**, *38* (6), 3098-3100.
11. Perdew, J. P., Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B* **1986**, *33* (12), 8822-8824.
12. Pantazis, D. A.; Neese, F., All-Electron Scalar Relativistic Basis Sets for the Lanthanides. *J. Chem. Theory Comput.* **2009**, *5* (9), 2229-2238.
13. Aravena, D.; Neese, F.; Pantazis, D. A., Improved Segmented All-Electron Relativistically Contracted Basis Sets for the Lanthanides. *J. Chem. Theory Comput.* **2016**, *12* (3), 1148-1156.
14. Lenthe, E. v.; Baerends, E. J.; Snijders, J. G., Relativistic regular two-component Hamiltonians. *J. Chem. Phys.* **1993**, *99* (6), 4597-4610.
15. Kendall, R. A.; Früchtl, H. A., The impact of the resolution of the identity approximate integral method on modern ab initio algorithm development. *Theor. Chem. Acc.* **1997**, *97* (1), 158-163.
16. Vahtras, O.; Almlöf, J.; Feyereisen, M. W., Integral approximations for LCAO-SCF calculations. *Chem. Phys. Lett.* **1993**, *213* (5), 514-518.
17. Takano, Y.; Houk, K. N., Benchmarking the Conductor-like Polarizable Continuum Model (CPCM) for Aqueous Solvation Free Energies of Neutral and Ionic Organic Molecules. *J. Chem. Theory Comput.* **2005**, *1* (1), 70-77.

3.4 A novel, rationally designed lanthanoid chelating tag delivers large paramagnetic structural restraints for biomolecular NMR

Reproduced from D. Joss, F. Winter, D. Häussinger, *Chemical Communications*, **2020**, 56, 12861-12864 with permission from the Royal Society of Chemistry.



Cite this: *Chem. Commun.*, 2020, 56, 12861

Received 22nd June 2020,
Accepted 10th September 2020

DOI: 10.1039/d0cc04337k

rsc.li/chemcomm

A novel, rationally designed lanthanoid chelating tag delivers large paramagnetic structural restraints for biomolecular NMR†

Daniel Joss, , Florine Winter and Daniel Häussinger *

Herein, a novel and rationally designed *ortho*-substituted pyridine activator is reported that reacts rapidly and selectively with cysteine thiols. It forms reduction-stable conjugates and induces large pseudocontact shifts, residual dipolar couplings and paramagnetic relaxation enhancement on both ubiquitin S57C and human carboxylate anhydrase II S50C constructs under physiological conditions.

Paramagnetic effects induced by metal centres with anisotropic distribution of unpaired electrons on biomacromolecules,^{1,2} e.g. lanthanoid ions complexed with suitable chelators for attachment, provide a unique and precise tool for the study of the properties of biomacromolecules by solution NMR.^{3–6} More specifically, lanthanoid chelating tags (LCT) induce long-range restraints that provide the unique opportunity to elucidate the structure of complexes, dynamics and interactions of biomacromolecules and ligands.^{7–16} In order to induce large paramagnetic effects, i.e. large pseudocontact shifts (PCS), residual dipolar couplings (RDC) and paramagnetic relaxation enhancement (PRE), rigid immobilization of the lanthanoid chelator on the surface of the biomacromolecule of interest is crucial.^{5,6} Conformational and positional invariant LCTs are ideally suited for monitoring protein interactions, as they can be placed far away from the interface and nevertheless provide accurate spatial information up to 200 Å.

Earlier approaches to suppress the translational and rotational freedom of the chelator on the surface of the protein, and hence to induce long-range restraints in solution, include: (1) two linker moieties attached to a double cysteine mutant,^{17–19} (2) restriction of the translational and rotational mobility,^{20–22} (3) placement of bulky groups on the macrocyclic ring scaffold.^{22,23}

In order to restrict the remaining rotational averaging by *para*-substituted linker moieties,^{20,21} we rationally designed a novel, *ortho*-substituted pyridine activator that leads to a

hampered rotation around the S_{cysteine}–C_{pyridine} bond and to a short, reduction-stable linkage between the LCT and the biomacromolecule (Fig. 1). As carved out during the design process, cysteine-selective and rapid conjugation in *ortho*-position requires a very delicate tuning of an electron-withdrawing π -acceptor type functional group and a suitable leaving group in order to balance high reactivity, potential hydrolysis and steric demands of the incoming cysteine thiol as nucleophile. During the nucleophilic aromatic substitution tagging reaction, a nitro substituent in *meta*-position stabilises the Meisenheimer complex intermediate²⁴ and, in combination with a methylsulphonyne leaving group, allows for fast and selective ligation of the LCT with a cysteine residue of the targeted biomacromolecule.

The seven methyl groups on the cyclen-based ring scaffold of the proposed LCT (Ln-(2*R*,2'*R*,2''*R*)-2,2',2''-((2*S*,5*S*,8*S*,11*S*)-2,5,8,11-tetramethyl-10-((6-(methylsulphonyl)-5-nitropyridin-2-yl)methyl)-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)triopropanoate = Ln-M7-Nitro), and on the pendant carboxylate side arms provide a significant rigidification of the chelator.^{25–28} Reduction-stable thioether linkages of LCT-protein were earlier shown to provide an attachment suitable for acquisition of paramagnetic restraints by in-cell NMR.^{12,13}

In order to demonstrate the applicability of our proposed approach, we synthesized paramagnetic thulium, dysprosium,

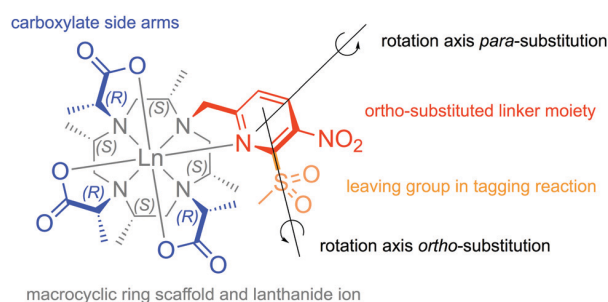


Fig. 1 Structure of Ln-M7-Nitro in $\Lambda(\delta\delta\delta\delta)$ conformation and indication of the important parts of the chelator.

Department of Chemistry, University of Basel, St. Johanns-Ring 19, Basel 4056, Switzerland. E-mail: daniel.haeussinger@unibas.ch

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d0cc04337k



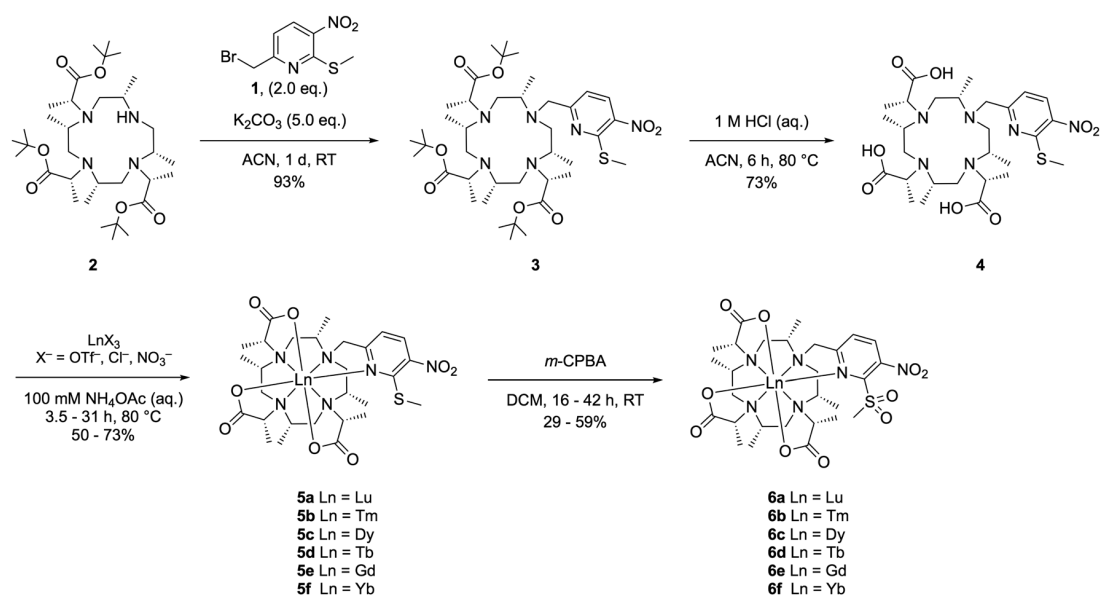


Fig. 2 Synthetic route to Ln-M7-Nitro starting from sevenfold methyl-substituted cyclen precursor **2** via alkylation, deprotection, metalation and oxidation.

terbium, ytterbium and gadolinium complexes in combination with the lutetium complex as diamagnetic reference and investigated their potential for the generation of PCSs (position of amide groups in 3D space), RDCs (orientation of the amide N-H vectors in space) and PRE (distance of amide protons from the paramagnetic centre) on both ubiquitin S57C (9 kDa) and human carbonic anhydrase II S50C (hCA II, 30 kDa) constructs.

With the idea of the installation of a leaving group in the *ortho*-position of the pyridine moiety of the envisioned LCT in mind, a suitable bromopyridine linker precursor **1** was synthesized starting from the commercially available 2-chloro-6-methyl-3-nitropyridine (ESI,† pp. 3–4). A sevenfold methyl-substituted cyclen precursor with three side arms already

installed (**2**) was then reacted with the brominated linker precursor **1** in an alkylation reaction (Fig. 2). The precursor of the LCT with all four side arms installed (**3**) was then deprotected by heating of the compound with hydrochloric acid in acetonitrile to give the free ligand **4**. Subsequently, the chelator was loaded with lanthanoid ions suitable for the investigation of paramagnetic effects on biomacromolecules (**5**) and the leaving group was activated by oxidation of the thioether to the methylsulphone (**6**).

In order to test the magnitude of the induced PCS and RDC effects of Ln-M7-Nitro on ubiquitin S57C and selectively ^{15}N leucine labelled hCA II S50C protein constructs, paramagnetic Tm-, Dy-, Tb- and Yb-loaded complexes were conjugated to the

Table 1 Induced axial and rhombic components of the paramagnetic susceptibility tensors ($\Delta\chi_{\text{ax}}$ and $\Delta\chi_{\text{rh}}$, in 10^{-32} m^3), metal position in PDB coordinate frame (X_{metal} , Y_{metal} , Z_{metal} , in Å), Euler angles (α , β , γ , in °) and quality factor (Q , mathematical definition in ESI, p. 1) on ubiquitin S57C (pH 6.0) and selectively ^{15}N leucine labelled hCA II S50C (pH 6.8) at 298 K. PCS data was fitted using Numbat,³² while RDC tensors were obtained with Fanten.³³ A detailed table with single and joint fits of PCS/RDC is available in the ESI (Tables S17 and S18)

Protein mutant	PDB	N° PCS	Ln ³⁺	$\Delta\chi_{\text{ax}}$ (10^{-32} m^3)	$\Delta\chi_{\text{rh}}$ (10^{-32} m^3)	X_{metal} (Å)	Y_{metal} (Å)	Z_{metal} (Å)	α (°)	β (°)	γ (°)	Q (%)
Ubiquitin S57C	1UBI ³⁴	32	Tm	60.9 ± 1.9	12.3 ± 1.3	16.3	16.9	12.3	110.9	70.4	101.4	5.1
		32	Dy	94.3 ± 2.2	10.7 ± 0.7				133.7	161.5	128.4	2.4
		44	Tb	65.9 ± 1.6	27.9 ± 0.8				128.0	162.4	110.4	7.0
hCA II S50C	3KS3 ³⁵	10	Tm	71.0 ± 1.8	6.2 ± 0.9	-31.0	4.4	14.9	174.7	17.4	139.0	4.3
		22	Dy	95.9 ± 1.1	13.1 ± 2.5				179.8	112.5	64.5	2.1
		38	Tb	63.6 ± 0.9	22.5 ± 1.4				4.9	62.5	109.5	2.0
		22	Yb	13.8 ± 0.6	7.5 ± 0.4				36.7	163.0	115.4	3.5
Protein mutant	PDB	N° RDC	Ln ³⁺	$\Delta\chi_{\text{ax}}$ (10^{-32} m^3)	$\Delta\chi_{\text{rh}}$ (10^{-32} m^3)	X_{metal} (Å)	Y_{metal} (Å)	Z_{metal} (Å)	α (°)	β (°)	γ (°)	Q (%)
Ubiquitin S57C	1UBI ³⁴	17	Tm	66.3	-13.1	16.3	16.9	12.3	164.3	-103.2	107.4	18.0
		15	Dy	106.7	-3.5				-131.4	-14.7	137.1	18.2
		21	Tb	79.2	-32.5				143.2	-15.7	146.2	20.5
hCA II S50C	3KS3 ³⁵	18	Tb	74.9	-23.3	-31.0	4.4	14.9	-23.2	-111.9	-3.3	18.3
		11	Yb	13.9	-6.0				-126.4	-151.1	-123.3	50.1



solvent exposed single cysteine residues on the protein constructs, whereas the Lu-complex was used as diamagnetic reference when conjugated to the protein. Subsequently, ^1H - ^{15}N HSQC (acquisition of PCSs) and ^1H - ^{15}N HSQC IPAP (acquisition of RDCs, no decoupling in the ^{15}N dimension) experiments were acquired. Ln-M7-Nitro was thereby shown to induce anisotropy parameters on ubiquitin S57C of up to $\Delta\chi_{\text{ax}} = 94.3 \times 10^{-32} \text{ m}^3$ for Dy-, $\Delta\chi_{\text{ax}} = 65.9 \times 10^{-32} \text{ m}^3$ for Tb- and $\Delta\chi_{\text{ax}} = 60.9 \times 10^{-32} \text{ m}^3$ for Tm-M7-Nitro (Table 1). When compared to other LCTs, the anisotropy parameters induced by Ln-M7-Nitro are exceedingly large and the value for the Dy-loaded chelator attached to ubiquitin S57C constitutes an unprecedented value of $\Delta\chi_{\text{ax}}$ for an LCT-protein construct.^{17,19,20,22,29} The obtained *Q*-factors and the distance of the fitted lanthanoid metal centre to the C_{beta} -carbon of the cysteine residue (6.9 Å) indicate excellent fits. Notably, RDCs up to 41.3 Hz were obtained for the Dy-loaded chelator, while RDCs of up to 35.9 Hz were obtained using the Tb complex (Tables S11 and S12, ESI[†]). In order to confirm the results obtained on ubiquitin S57C (9 kDa) and to extend our methodology to a larger protein, selectively ^{15}N leucine labelled hCA II S50C (30 kDa) with a single solvent exposed cysteine was chosen as suitable target. Similar anisotropy parameters as for ubiquitin S57C were found (Table 1) and the Ln- $\text{C}_{\text{beta,cys}}$ distance of 7.1 Å indicated an excellent quality of the obtained fits. As can be seen from the anisotropy parameters depicted as isosurfaces (Fig. 3), the properties of the induced effects match for the pairs Dy and Tb as well as for Tm and Yb – intrinsic properties of the lanthanoids that were described in literature.⁵ When compared to seven-fold methyl substituted, pyridine derived LCTs, as e.g. Tm-M7FPy-DOTA²¹ attached to ubiquitin S57C, Tm-M7-Nitro shows a superior performance, which can most likely be attributed to the switch of the ligation site on the pyridine linker moiety from the *para*- to the *ortho*-position and therefore to an increased rotational and translational rigidity of

the LCT relative to the protein surface. Since PCSs and RDCs are susceptible to different motional averaging mechanisms, a joint fit of PCSs and RDCs should result in the same tensors as the individual fits in case of a non-mobile LCT.^{30,31} We were delighted to find that $\Delta\chi_{\text{ax}}$ for the combined PCS/RDC fit agreed on average with the PCS fit within 5 and 9% for ubiquitin and hCA II, respectively, whereas the agreement with the RDC fit was even better (0.3 and 1%, see Table S18, ESI[†]). Ln-M7-Nitro might, therefore, be also useful to detect dynamics. The moderate Curie line broadening of approx. only 20 Hz in the case of hCA II, observed at 600 MHz in 24 Å distance from the lanthanoid, holds great promise also for applications at ultra-high fields (Fig. S40, ESI[†]).

Due to the stabilization of the intermediate Meisenheimer complex²⁴ during the ligation reaction by the newly developed activator, Ln-M7-Nitro provides a fast, quantitative and therefore clean ligation to the protein. As confirmed by ESI-HRMS, a ligation yield of > 95% is already observed after 30 min at room temperature (Fig. 4). The 3-nitro-substituted 2-(methylsulphonyl)pyridine constitutes a novel, highly reactive and selective moiety for tagging of cysteines in biomacromolecules under close to physiological conditions (highly diluted (~100 μM), aqueous buffer, pH 7.0). While the methylsulphonyl leaving group shows a suitable balance in terms of reactivity and selectivity, a fluorine leaving group led to hydrolysis and chlorine showed no reaction.

Besides the favourable properties of Ln-M7-Nitro for the generation of anisotropic paramagnetic effects, we were curious about the properties of Ln-M7-Nitro as PRE tag. PRE is an isotropic effect that enhances the relaxation of spins in vicinity to the paramagnetic metal centre and causes thereby line broadening, which can be quantified by plotting the intensity ratio of paramagnetic and diamagnetic peaks *vs.* the distance of the amide proton to the metal centre. Upon attachment of the Gd complex of the newly developed LCT to both ubiquitin S57C

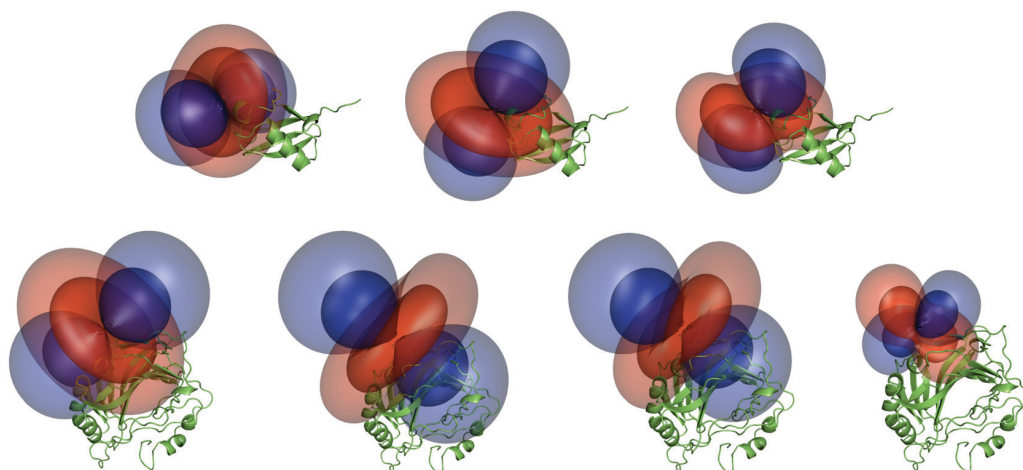


Fig. 3 Anisotropy parameters induced by Ln-M7-Nitro on ubiquitin S57C (top) and selectively ^{15}N leucine labelled hCA II S50C (bottom) depicted as isosurfaces with a given shift in ppm (generated using Numbat³²). Top (from left to right): Tm-, Dy- and Tb-M7-Nitro-Ub^{S57C} (inner layer: 4.0 ppm, outer layer: 1.5 ppm). Bottom (from left to right): Tm-, Dy-, Tb- and Yb-M7-Nitro attached to selectively ^{15}N leucine labelled hCA II S50C (inner layer: 3.0 ppm, outer layer: 0.8 ppm).



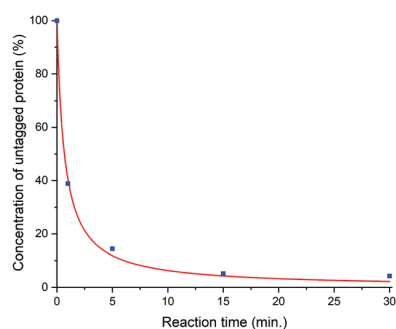


Fig. 4 Kinetics of the tagging reaction of monomeric ubiquitin S57C (90 μM) with Lu-M7-Nitro (540 μM) in phosphate buffer (10 mM, pH 7.0, 0.2 mM TCEP) followed by ESI-HRMS. The experimental data points were fitted with the equation for second order kinetics $[\text{Ubi}] = (k \cdot t + 1/[\text{Ubi}]_{t=0})^{-1}$ to give $k = 274.5 \text{ M}^{-1} \text{ s}^{-1}$. A comparison of Ln-M7-Nitro and previously reported LCTs is shown in Fig. S41 (ESI †).

and hCA II S50C, strong PREs were detected in ^1H - ^{15}N HSQC experiments and evaluated with respect to their distance dependence, the amino acid sequence of the protein and the theoretically predicted values for an LCT-protein construct with the same molecular weight (Fig. S9–S12, ESI †).³⁶ PRE was previously used to resolve structures and interactions of biomacromolecules in solution.³⁷

To conclude, a novel, rationally designed *ortho*-substituted pyridine activator was developed that delivers an LCT with unprecedented paramagnetic properties and, thereby, allows for acquisition of experimental structural restraints for the investigation of structure and dynamics of biomacromolecules and their complexes *via* PCSs, RDCs and PREs. The fast and clean ligation reaction of the new activator in combination with the strong, induced paramagnetic effects render Ln-M7-Nitro as an ideal LCT for challenging applications.

The Chemistry Department of the University of Basel and the SNSF grant 200021_130263 are acknowledged for financial support. Biological structures were generated using the open source software PyMOL. C. E. Housecroft, E. C. Constable, T. Müntener, R. Vogel and P. Rieder are acknowledged for helpful discussions and S. Fischer for involvement in synthetic side projects.

Conflicts of interest

There are no conflicts to declare.

Notes and references

- I. Bertini, C. Luchinat, G. Parigi and R. Pierattelli, *ChemBioChem*, 2005, **6**, 1536–1549.
- J. H. Prestegard, H. M. Al-Hashimi and J. R. Tolman, *Q. Rev. Biophys.*, 2000, **33**, 371–424.
- J. Koehler and J. Meiler, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2011, **59**, 360–389.
- W.-M. Liu, M. Overhand and M. Ubbink, *Coord. Chem. Rev.*, 2014, **273–274**, 2–12.
- C. Nitsche and G. Otting, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2017, **98–99**, 20–49.
- D. Joss and D. Häussinger, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2019, **114–115**, 284–312.
- G. Pintacuda, M. John, X.-C. Su and G. Otting, *Acc. Chem. Res.*, 2007, **40**, 206–212.
- J.-Y. Guan, P. H. J. Keizers, W.-M. Liu, F. Löhr, S. P. Skinner, E. A. Heeneman, H. Schwalbe, M. Ubbink and G. Siegal, *J. Am. Chem. Soc.*, 2013, **135**, 5859–5868.
- C. Göbl, T. Madl, B. Simon and M. Sattler, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2014, **80**, 26–63.
- K. D. Brewer, T. Bacaj, A. Cavalli, C. Camilloni, J. D. Swarbrick, J. Liu, A. Zhou, P. Zhou, N. Barlow, J. Xu, A. B. Seven, E. A. Prinslow, R. Voleti, D. Häussinger, A. M. J. J. Bonvin, D. R. Tomchick, M. Vendruscolo, B. Graham, T. C. Südhof and J. Rizo, *Nat. Struct. Mol. Biol.*, 2015, **22**, 555–564.
- T. Saio, K. Ogura, H. Kumeta, Y. Kobashigawa, K. Shimizu, M. Yokochi, K. Kodama, H. Yamaguchi, H. Tsujishita and F. Inagaki, *Sci. Rep.*, 2015, **5**, 16685.
- T. Müntener, D. Häussinger, P. Selenko and F.-X. Theillet, *J. Phys. Chem. Lett.*, 2016, **7**, 2821–2825.
- B.-B. Pan, F. Yang, Y. Ye, Q. Wu, C. Li, T. Huber and X.-C. Su, *Chem. Commun.*, 2016, **52**, 10237–10240.
- K. Zimmermann, D. Joss, T. Müntener, E. S. Nogueira, M. Schäfer, L. Knörr, F. W. Monnard and D. Häussinger, *Chem. Sci.*, 2019, **10**, 5064–5072.
- D. Joss, R. Vogel and D. Häussinger, *Ref. Mod. in Chem, Mol. Sci. and Chem. Eng.*, 2020, DOI: 10.1016/B978-0-12-409547-2.14848-6.
- C. A. Barnes, Y. Shen, J. Ying, Y. Takagi, D. A. Torchia, J. R. Sellers and A. Bax, *J. Am. Chem. Soc.*, 2019, **141**, 9004–9017.
- P. H. Keizers, A. Saragliadis, Y. Hiruma, M. Overhand and M. Ubbink, *J. Am. Chem. Soc.*, 2008, **130**, 14802–14812.
- W.-M. Liu, P. H. J. Keizers, M. A. S. Hass, A. Blok, M. Timmer, A. J. C. Sarris, M. Overhand and M. Ubbink, *J. Am. Chem. Soc.*, 2012, **134**, 17306–17313.
- M. D. Lee, M. L. Dennis, B. Graham and J. D. Swarbrick, *Chem. Commun.*, 2017, **53**, 13205–13208.
- F. Yang, X. Wang, B.-B. Pan and X.-C. Su, *Chem. Commun.*, 2016, **52**, 11535–11538.
- T. Müntener, J. Kottelat, A. Huber and D. Häussinger, *Bioconjugate Chem.*, 2018, **29**, 3344–3351.
- D. Joss and D. Häussinger, *Chem. Commun.*, 2019, **55**, 10543–10546.
- D. Joss, M.-S. Bertrams and D. Häussinger, *Chem. – Eur. J.*, 2019, **25**, 11910–11917.
- C. N. Neumann, J. M. Hooker and T. Ritter, *Nature*, 2016, **534**, 369–373.
- D. Häussinger, J.-R. Huang and S. Grzesiek, *J. Am. Chem. Soc.*, 2009, **131**, 14761–14767.
- A. C. L. Opina, M. Strickland, Y.-S. Lee, N. Tjandra, R. A. Byrd, R. E. Swenson and O. Vaslatiy, *Dalton Trans.*, 2016, **45**, 4673–4687.
- D. Joss, R. M. Walliser, K. Zimmermann and D. Häussinger, *J. Biomol. NMR*, 2018, **29**, 29–38.
- D. Parker, R. S. Dickins, H. Puschmann, C. Crossland and J. A. K. Howard, *Chem. Rev.*, 2002, **102**, 1977–2010.
- F. Peters, M. Maestre-Martinez, A. Leonov, L. Kovacic, S. Becker, R. Boelens and C. Griesinger, *J. Biomol. NMR*, 2011, **51**, 329–337.
- D. Shishmarev and G. Otting, *J. Biomol. NMR*, 2013, **56**, 203–216.
- I. Bertini, C. Luchinat, G. Parigi and E. Ravera, *NMR of Paramagnetic Molecules*, Elsevier, Boston, 2nd edn, 2017, pp. 277–312.
- C. Schmitz, M. J. Stanton-Cook, X.-C. Su, G. Otting and T. Huber, *J. Biomol. NMR*, 2008, **41**, 179.
- M. Rinaldelli, A. Carlon, E. Ravera, G. Parigi and C. Luchinat, *J. Biomol. NMR*, 2015, **61**, 21–34.
- R. Ramage, J. Green, T. W. Muir, O. M. Ogunjobi, S. Love and K. Shaw, *Biochem. J.*, 1994, **299**(Pt 1), 151–158.
- B. S. Avaru, C. U. Kim, K. H. Sippel, S. M. Gruner, M. Agbandje-McKenna, D. N. Silverman and R. McKenna, *Biochemistry*, 2010, **49**, 249–251.
- J. L. Battiste and G. Wagner, *Biochemistry*, 2000, **39**, 5355–5365.
- G. M. Clore and J. Iwahara, *Chem. Rev.*, 2009, **109**, 4108–4139.



**A novel, rationally designed lanthanoid chelating tag delivers large
paramagnetic structural restraints for biomolecular NMR**

Supporting Information

Daniel Joss,^a Florine Winter,^a and Daniel Häussinger*^a

^a Department of Chemistry, University of Basel, St. Johanns-Ring 19, CH-4056 Basel, daniel.haeussinger@unibas.ch

General Remarks

Unless otherwise stated, reactions were performed under an argon atmosphere and chemicals were used as received without further purification. All reactions were executed in anhydrous solvents, while technical grade solvents were used for extractions and flash column chromatography.

Silica gel 60 F₂₅₄ on aluminium sheets from Merck was used for thin layer chromatography. The detection was either monitored with an UV-lamp at a wavelength of 254 nm or with potassium permanganate stain.

NMR experiments were performed at a temperature of 298 K on Bruker Avance III NMR spectrometers operating at 400, 500 and 600 MHz. ESI-MS spectra were recorded on a Shimadzu LCMS-2020 liquid chromatograph mass spectrometer. HRMS spectra were measured on a Bruker MaXis 4G HiRes ESI Mass Spectrometer.

Fitting of kinetic data and PRE analysis were performed using OriginPro 2018 (OriginLab, Northampton, MA, United States). Theoretical PRE were calculated using the reduced Solomon-Bloembergen equation with $J=7/2$ for Gd and rotational correlation times for both protein constructs approximated with $0.6 \times$ molecular mass of the LCT-protein constructs (5.7 ns for ubiquitin S57C labelled with Gd-M7-Nitro and 18.1 ns for hCA S50C labelled with Gd-M7-Nitro).¹ I_{para}/I_{dia} ratios were normalized using the I_{para}/I_{dia} ratio of the most distant residue. Tri-*tert*-butyl 2,2',2''-((2S,5S,8S,11S)-2,5,8,11-tetramethyl-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)-(2R,2'R,2''R)-tripropionate was synthesized according to the published procedures by Müntener et al.² Ubiquitin S57C was expressed as described previously by Sass et al.,³ selectively ¹⁵N leucine labelled hCA II S50C as described by Varghese et al.⁴ Tagging reactions were performed in 10 mM phosphate and 0.2 mM TCEP buffer with pH 7.0 (ubiquitin S57C) or pH 6.8 (hCA S50C) at rt overnight. ¹H-¹⁵N HSQC and ¹H-¹⁵N HSQC IPAP spectra were measured in 10 mM phosphate buffer with pH 6.0 (ubiquitin S57C, ca. 100 μM) and pH 6.8 (hCA II S50C, ca. 200 μM) at a temperature of 298 K on a 600 MHz Bruker Avance III NMR spectrometer equipped with a cryogenic QCI-F probe. The obtained NMR spectra were assigned using CcpNmr Analysis.⁵ The tensor properties were then obtained by fitting the residues in secondary structure elements of ubiquitin (PDB 1UBI)⁶ or the leucine residues of hCA II (PDB 3KS3)⁷ using Nubat (PCSs) with correction for residual anisotropic chemical shifts (RACS),⁸ while RDCs were fitted using Fanten (with order parameter $S^2 = 0.9$) and Paramagpy.⁹⁻¹⁰ Absolute deviation values of axial and rhombic components of PCS-derived tensors using Nubat were estimated by a Monte-Carlo simulation with a perturbation level of 0.1 Å on the protein structure and 10000 iterations. The metal centres were found in a distance of 6.9 Å (ubiquitin S57C) and 7.1 Å (hCA II S50C) from the C_{beta} of the cysteine residue. Q-factors were calculated using the following equation:

$$Q = \frac{\sqrt{\sum (PCS_{exp} - PCS_{calc})^2}}{\sqrt{\sum (PCS_{exp})^2}}$$

HPLC conditions

Analytical HPLC measurements were performed on a Shimadzu LC system (LC-20AT prominence liquid chromatograph, SIL-20A HT prominence auto sampler, CTO-20AC prominence column oven, CBM-20A communications bus module, SPD-20A prominence UV/VIS detector ($\lambda = 254$ and 280 nm), LC-20AD prominence liquid chromatograph) combined with a Shimadzu LCMS-2020 liquid chromatograph mass spectrometer. As column for analytical HPLC measurements, a ReproSil-Pur ODS, $3.3 \mu\text{m}$, 150×3 mm, provided by Maisch GmbH was used. Commercial HPLC grade solvents were used and a binary gradient was applied.

Solvent A: Milli-Q water + 0.1% TFA

Solvent B: 90% acetonitrile + 10% Milli-Q water + 0.085% TFA.

HPLC gradient: 95% A (min 0-2), linear gradient 95% A to 100% B (min 2-6), 100% B (min 6-14), linear gradient 100% B to 95% A (min 14-15), 95% A (min 15-22).

Semi-preparative HPLC purification was performed on a Shimadzu LC system (LC-20AT prominence liquid chromatograph, SIL-20A HT prominence auto sampler, CTO-20AC prominence column oven, CBM-20A communications bus module, SPD-20A prominence UV/VIS detector ($\lambda = 254$ and 280 nm), LC-20AD prominence liquid chromatograph) combined with a Shimadzu LCMS-2020 liquid chromatograph mass spectrometer. As column for preparative HPLC purification, a ReproSil-Pur 120 ODS-3, $5 \mu\text{m}$, 150×20 mm, provided by Maisch GmbH was used. Commercial HPLC grade solvents were used and a binary gradient was applied during purification.

Solvent A: Milli-Q water + 0.1% TFA

Solvent B: 90% acetonitrile + 10% Milli-Q water + 0.085% TFA.

HPLC gradient: 95% A (min 0-2), linear gradient 95% A to 100% B (min 2-15), 100% B (min 15-22), linear gradient 100% B to 95% A (min 22-23), 95% A (min 23-25).

HPLC measurements of protein samples were performed using the direct injection mode on a Shimadzu LC system (LC-20AT prominence liquid chromatograph, SIL-20A HT prominence auto sampler, CTO-20AC prominence column oven, CBM-20A communications bus module, SPD-20A prominence UV/VIS detector ($\lambda = 254$ and 280 nm), LC-20AD prominence liquid chromatograph) combined with a Shimadzu LCMS-2020 liquid chromatograph mass spectrometer. Commercial HPLC grade solvents were used and a binary gradient was applied. MS spectra of proteins were deconvoluted using the Bruker Daltonics DataAnalysis software.

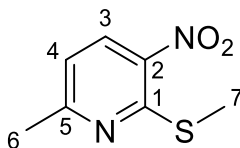
Solvent A: Milli-Q water + 0.1% TFA

Solvent B: 90% acetonitrile + 10% Milli-Q water + 0.085% TFA.

HPLC gradient: isocratic 95% A (min 0-4).

6-Methyl-2-(methylthio)-3-nitropyridine

Sodium methanethiolate (487 mg, 6.95 mmol, 1.2 eq.) was added to a solution of 2-chloro-6-methyl-3-nitropyridine (1.00 g, 5.79 mmol, 1.0 eq.) in DMF (20 mL) at 0 °C and the mixture was stirred for 2.5 h at rt. The reaction mixture was then concentrated under reduced pressure and the crude was purified by flash column chromatography (SiO₂, cyclohexane:EtOAc, 9:1) to obtain 6-(bromomethyl)-2-(methylthio)-3-nitropyridine (760 mg, 4.13 mmol, 71%) as yellow solid.



TLC (SiO₂, cyclohexane:EtOAc, 9:1): R_f = 0.27.

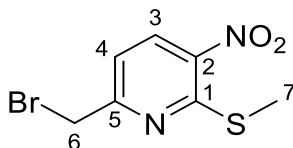
¹H-NMR (500 MHz, CDCl₃, δ/ppm): 8.38 (d, ³J_{H3-H4} = 8.3 Hz, 1H, **H₃**), 7.00 (d, ³J_{H4-H3} = 8.3 Hz, 1H, **H₄**), 2.63 (s, 3H, **H₆**), 2.57 (s, 3H, **H₇**).

¹³C-NMR (126 MHz, CDCl₃, δ/ppm): 163.7 (**C₅**), 158.0 (**C₁**), 140.3 (**C₂**), 133.9 (**C₃**), 118.2 (**C₄**), 25.0 (**C₆**), 14.4 (**C₇**).

HRMS: [M+H]⁺ C₇H₉N₂O₂S, m/z (calc.) = 185.0379, m/z (meas.) 185.0378.

6-(Bromomethyl)-2-(methylthio)-3-nitropyridine

N-Bromosuccinimide (930 mg, 5.23 mmol, 1.3 eq.) followed by dibenzoyl peroxide (130 mg, 0.402 mmol, 0.1 eq., 75%) was added to a solution of 6-methyl-2-(methylthio)-3-nitropyridine (741 mg, 4.02 mmol, 1.0 eq.) in dichloroethane (10 mL). The reaction mixture was heated to 80 °C and NBS (358 mg, 2.01 mmol, 0.5 eq.) and dibenzoyl peroxide (390 mg, 1.21 mmol, 0.3 eq., 75%) were added within 6 h. After 22 h, the reaction mixture was poured onto aq. sat. NaHCO₃ solution and the aq. layer was extracted with dichloromethane (3 x 10 ml). The combined organic layers were washed with brine, dried over anhydrous Na₂SO₄, filtered and concentrated *in vacuo*. The crude product was purified by flash column chromatography (SiO₂, cyclohexane:EtOAc, 19:1) to yield 6-(bromomethyl)-2-(methylthio)-3-nitropyridine (240 mg, 0.913 mmol, 23%) as a yellow solid.



TLC (SiO₂, cyclohexane:EtOAc, 19:1): R_f = 0.16.

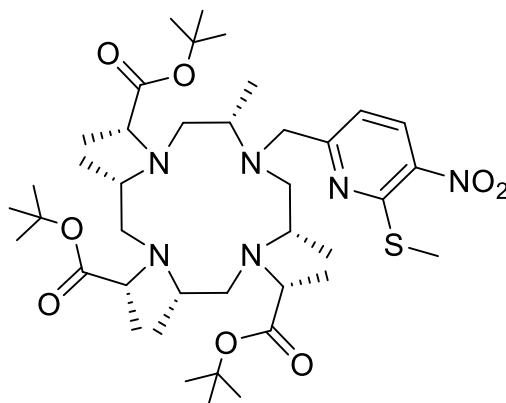
¹H-NMR (400 MHz, CDCl₃, δ/ppm): 8.49 (d, ³J_{H3-H4} = 8.3 Hz, 1H, **H₃**), 7.30 (d, ³J_{H4-H3} = 8.4 Hz, 1H, **H₄**), 4.55 (s, 2H, **H₆**), 2.59 (s, 3H, **H₇**).

¹³C-NMR (126 MHz, CDCl₃, δ/ppm): 160.9 (**C₅**), 158.6 (**C₁**), 141.1 (**C₂**), 134.8 (**C₃**), 118.1 (**C₄**), 32.2 (**C₆**), 14.5 (**C₇**).

HRMS: [M+H]⁺ C₇H₈BrN₂O₂S, m/z (calc.) = 262.9484, m/z (meas.) 262.9483.

Tri-*tert*-butyl 2,2',2''-((2S,5S,8S,11S)-2,5,8,11-tetramethyl-10-((6-(methylthio)-nitropyridin-2-yl)methyl)-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)-(2R,2'R,2''R)-tripropionate

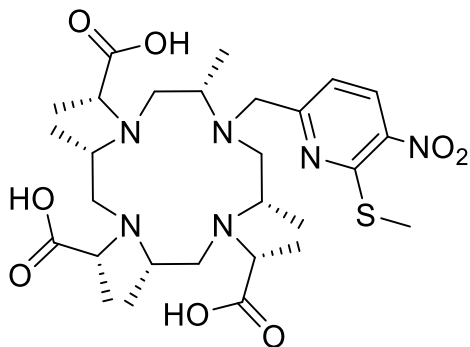
A solution of 6-(bromomethyl)-2-(methylthio)-3-nitropyridine (42.9 mg, 163 μmol , 2.0 eq.) in dry acetonitrile (1.0 mL) was added under argon to a suspension of tri-*tert*-butyl 2,2',2''-((2S,5S,8S,11S)-2,5,8,11-tetramethyl-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)-(2R,2'R,2''R)-tripropionate (50.0 mg, 81.6 μmol , 1.0 eq.) and potassium carbonate (56.4 mg, 204 μmol , 5.0 eq.) in dry acetonitrile (1.0 mL). The reaction mixture was stirred for 24 h at rt. The suspension was filtered and the crude product was purified by prep. HPLC to obtain tri-*tert*-butyl 2,2',2''-((2S,5S,8S,11S)-2,5,8,11-tetramethyl-10-((6-(methylthio)-nitropyridin-2-yl)methyl)-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)-(2R,2'R,2''R)-tripropionate (60.0 mg, 75.5 μmol , 93%) as a yellow oil.



HRMS: $[\text{M}+\text{H}]^+$ $\text{C}_{40}\text{H}_{71}\text{N}_6\text{O}_8\text{S}$, m/z (calc.) = 795.5049, m/z (meas.) 795.5048.

(2*R*,2'*R*,2''*R*)-2,2',2''-((2*S*,5*S*,8*S*,11*S*)-2,5,8,11-tetramethyl-10-((6-(methylthio)-5-nitropyridin-2-yl)methyl)-1,4,7,10-tetraazacyclododecane-1,4,7-triyl) tripropionic acid

Tri-*tert*-butyl 2,2',2''-((2*S*,5*S*,8*S*,11*S*)-2,5,8,11-tetramethyl-10-((6-(methylthio)-5-nitropyridin-2-yl)methyl)-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)-(2*R*,2'*R*,2''*R*)-tripropionate (60.0 mg, 75.5 μ mol, 1.0 eq.) was dissolved in acetonitrile (1.0 mL) and aq. HCl (1 M, 2.0 mL). The resulting solution was heated to 80 °C for 6 h. Then the mixture was cooled to rt and purified by prep. HPLC to yield (2*R*,2'*R*,2''*R*)-2,2',2''-((2*S*,5*S*,8*S*,11*S*)-2,5,8,11-tetramethyl-10-((6-(methylthio)-5-nitropyridin-2-yl)methyl)-1,4,7,10-tetraazacyclododecane-1,4,7-triyl) tripropionic acid (34.5 mg, 55.0 μ mol, 73%) as a white solid.



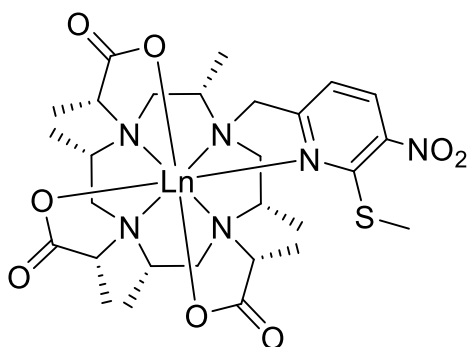
HRMS: [M+H]⁺ C₂₈H₄₇N₆O₈S, m/z (calc.) = 627.3171, m/z (meas.) 627.3175.

Ln-(2*R,2'R,2''R*)-2,2',2''-((2*S,5*S,8*S,11*S)-2,5,8,11-tetramethyl-10-((6-(methylthio)-5-nitropyridin-2-yl)methyl)-1,4,7,10-tetraazacyclododecane-1,4,7-triyl) tripropionate, Ln-M7-Nitro-SMe****

LnX₃ (4.0 eq.) was added to a solution of (2*R,2'R,2''R*)-2,2',2''-((2*S,5*S,8*S,11*S**)-2,5,8,11-tetramethyl-10-((6-(methylthio)-5-nitropyridin-2-yl)methyl)-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)-(2*R,2'R,2''R*)-tripropionic acid (1.0 eq.) in aq. ammonium acetate (100 mM). The reaction mixture was stirred in a pre-heated oil-bath at 80 °C. The reaction mixture was cooled to rt and purified by prep. HPLC to obtain the title compound **50a-e** as an off-white solid.**

Table S1: Reaction conditions for the synthesis of Ln-M7-Nitro-SMe.

	Free Ligand	aq. CH ₃ COONH ₄	LnX ₃	Time	Yield
Lu	12.5 mg, 19.9 μmol	6 mL	Lu(OTf) ₃ (49.8 mg, 80.0 μmol)	16 h	11.5 mg, 14.4 μmol, 72%
Tm	10.0 mg, 16.0 μmol	10 mL	Tm(OTf) ₃ (19.7 mg, 31.9 μmol)	3.5 h	8.2 mg, 10.4 μmol, 65%
Dy	8.0 mg, 12.8 μmol	8 mL	Dy(OTf) ₃ (31.2 mg, 51.2 μmol)	31 h	6.9 mg, 8.8 μmol, 69%
Tb	8.0 mg, 12.8 μmol	8 mL	TbCl ₃ · 6 H ₂ O (19.1 mg, 51.2 μmol)	31 h	6.7 mg, 8.6 μmol, 67%
Gd	8.0 mg, 12.8 μmol	8 mL	Gd(OTf) ₃ (30.9 mg, 51.2 μmol)	22 h	5.0 mg, 6.4 μmol, 50%
Yb	21.0 mg, 33.5 μmol	10 mL	Yb(NO ₃) ₃ · 5 H ₂ O (60.4 mg, 134.0 μmol)	16 h	19.5 mg, 24.5 μmol, 73%



Lu: HRMS: [M+H]⁺ C₂₈H₄₄LuN₆O₈S, m/z (calc.) = 799.2344, m/z (meas.) = 799.2343.

Tm: HRMS: [M+H]⁺ C₂₈H₄₄N₆O₈STm, m/z (calc.) = 793.2278, m/z (meas.) = 793.2270.

Dy: HRMS: [M+H]⁺ C₂₈H₄₄DyN₆O₈S, m/z (calc.) = 788.2228, m/z (meas.) = 788.2225.

Tb: HRMS: [M+H]⁺ C₂₈H₄₄N₆O₈STb, m/z (calc.) = 783.2189, m/z (meas.) = 783.2185.

Gd: HRMS: [M+H]⁺ C₂₈H₄₄GdN₆O₈S, m/z (calc.) = 782.2181, m/z (meas.) = 782.2179.

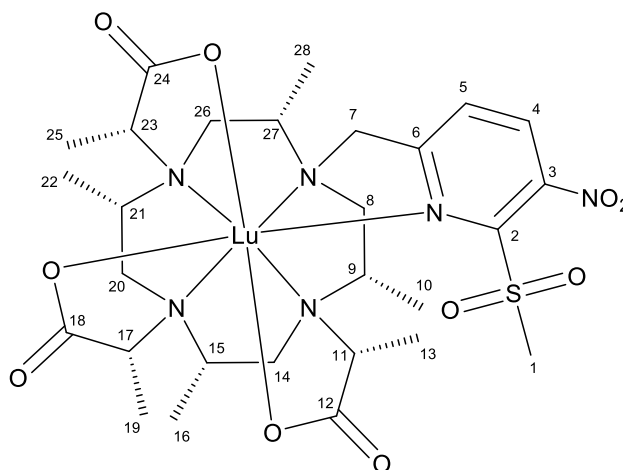
Yb: HRMS: [M+Na]⁺ C₂₈H₄₃NaN₆O₈SYb, m/z (calc.) = 820.2148, m/z (meas.) = 820.2145.

Ln-(2*R*,2'*R*,2''*R*)-2,2',2''-((2*S*,5*S*,8*S*,11*S*)-2,5,8,11-tetramethyl-10-((6-(methylsulphonyl)-5-nitropyridin-2-yl)methyl)-1,4,7,10-tetraazacyclododecane-1,4,7-triyl) tripropanoate, Ln-M7-Nitro

Meta-chloroperoxybenzoic acid (77%) was added to a solution of Lu-(2*R*,2'*R*,2''*R*)-2,2',2''-((2*S*,5*S*,8*S*,11*S*)-2,5,8,11-tetramethyl-10-((6-(methylthio)-5-nitropyridin-2-yl)methyl)-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)-(2*R*,2'*R*,2''*R*)-tripropanoate (1.0 eq.) in dichloromethane (1.0 - 5.0 mL). The resulting reaction mixture was stirred at rt. The reaction mixture was diluted with acetonitrile and dichloromethane was removed under reduced pressure. The remaining solution was purified by prep. HPLC to yield Ln-(2*R*,2'*R*,2''*R*)-2,2',2''-((2*S*,5*S*,8*S*,11*S*)-2,5,8,11-tetramethyl-10-((6-(methylsulphonyl)-5-nitropyridin-2-yl)methyl)-1,4,7,10-tetraazacyclododecane-1,4,7-triyl) tripropanoate as an off-white solid.

Table S2: Reaction conditions for the synthesis of Ln-M7-Nitro.

	Starting material	<i>m</i> -CPBA	Time	Yield
Lu	6.5 mg, 8.1 μmol	9.1 mg, 40.7 μmol	19 h	4.0 mg, 4.8 μmol, 59%
Tm	4.0 mg, 5.1 μmol	5.7 mg, 25.3 μmol	22 h	2.0 mg, 2.4 μmol, 48%
Dy	14.0 mg, 17.8 μmol	39.9 mg, 178 μmol	16 h	4.5 mg, 2.4 μmol, 31%
Tb	13.0 mg, 16.6 μmol	37.2 mg, 166 μmol	16 h	5.0 mg, 6.1 μmol, 37%
Gd	5.0 mg, 6.4 μmol	7.2 mg, 32.0 μmol	42 h	1.5 mg, 1.9 μmol, 29%
Yb	19.5 mg, 24.5 μmol	110.0 mg, 490 μmol	16 h	10.0 mg, 12.1 μmol, 49%



NMR-Assignment for Lu-M7-Nitro:

¹H-NMR (600 MHz, D₂O, δ/ppm): 8.56 (bs, 1H, **H₄**), 8.06 (d, ³*J*_{H5-H4} = 8.3 Hz, 1H, **H₅**), 4.48 (d, ²*J* = 14.4 Hz, 1H, **H_{7a}**), 4.43 (d, ²*J* = 14.1 Hz, 1H, **H_{7b}**), 4.10 – 3.90 (m, 2H, **H₁₁**, **H₉**), 3.76 (q, ³*J*_{H17-H19} = 7.3 Hz, 1H, **H₁₇**), 3.71 (q, ³*J*_{H23-H25} = 7.2 Hz, 1H, **H₂₃**), 3.60 (s, 3H, **H₁**), 3.24 – 3.14 (m, 2H, **H₂₇**, **H_{8eq}**), 3.06 – 3.03 (m, 2H, **H_{14ax}**, **H₁₅**), 3.03 – 2.98 (m, 2H, **H_{26ax}**, **H₂₁**), 2.98 – 2.92 (m, 1H, **H_{20ax}**), 2.79 (d, ²*J*_{H8ax-H8eq} = 15.3 Hz, 1H, **H_{8ax}**), 2.74 (d, ²*J*_{H14eq-H14ax} = 12.2 Hz, 1H, **H_{14eq}**), 2.69 – 2.64 (m, 2H, **H_{26eq}**, **H_{20eq}**), 1.52 (d, ³*J*_{H13-H11} = 7.1 Hz, 3H, **H₁₃**), 1.44 (d, ³*J*_{H19-H17} = 7.4 Hz, 3H, **H₁₉**), 1.42 (d, ³*J*_{H23-H25} = 7.0 Hz, 3H, **H₂₅**), 1.30 (d, ³*J*_{H10-H9} = 6.5 Hz, 3H, **H₁₀**), 1.17 (d, ³*J*_{H16-H15} = 5.6 Hz, 3H, **H₁₆**), 1.16 (d, ³*J*_{H22-H21} = 6.2 Hz, 3H, **H₂₂**), 0.54 (bs, 3H, **H₂₈**).

¹³C-NMR (151 MHz, D₂O, δ/ppm, extracted from HSQC and HMBC): 183.4 (**C₁₈**), 183.1 (**C₂₄**), 182.1 (**C₁₂**), 159.0 (**C₆**), 144.3 (**C₃**), 140.4 (**C₂**), 135.7 (**C₄**), 131.9 (**C₅**), 67.4 (**C₁₇**), 66.4 (**C₁₁**), 66.3 (**C₂₃**), 60.5 (**C₂₁**), 60.3 (**C₁₅**), 58.8 (**C₉**), 57.0 (**C₇**), 55.1 (**C₂₇**), 48.8 (**C₈**), 46.3 (**C₂₇**), 45.5 (**C₁₄**), 45.0 (**C₂₆**), 41.2 (**C₁**), 13.2 (**C₁₉**), 13.2 (**C₁₆**), 13.1 (**C₁₃**), 13.0 (**C₂₅**), 12.9 (**C₁₀**), 12.8 (**C₂₂**), 12.8 (**C₂₈**).

Lu: **HRMS:** [M+H]⁺ C₂₈H₄₄LuN₆O₁₀S, m/z (calc.) = 831.2242, m/z (meas.) = 831.2231.

Tm: **HRMS:** [M+H]⁺ C₂₈H₄₄N₆O₁₀STm, m/z (calc.) = 825.2176, m/z (meas.) = 825.2187.

Dy: **HRMS:** [M+H]⁺ C₂₈H₄₄DyN₆O₁₀S, m/z (calc.) = 820.2126, m/z (meas.) = 820.2118.

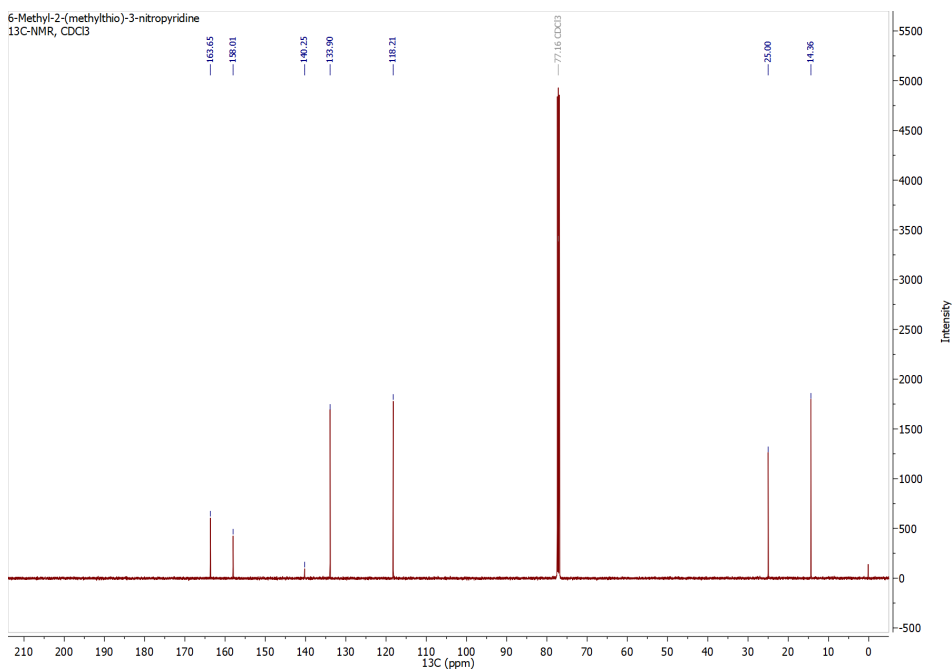
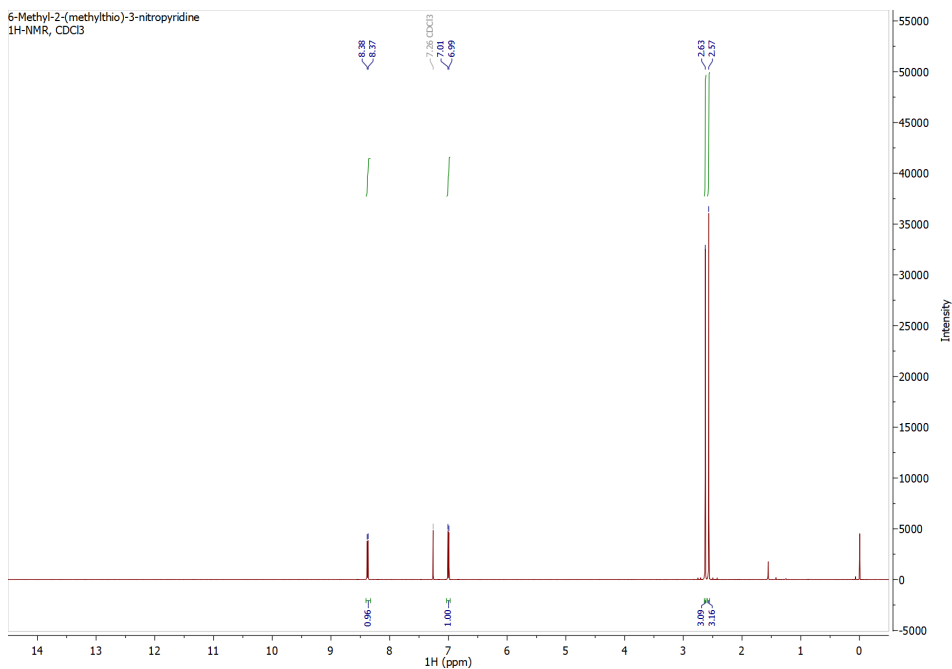
Tb: **HRMS:** [M+H]⁺ C₂₈H₄₄N₆O₁₀STb, m/z (calc.) = 815.2088, m/z (meas.) = 815.2079.

Gd: HRMS: $[M+H]^+$ C₂₈H₄₄GdN₆O₁₀S, m/z (calc.) = 814.2080, m/z (meas.) = 814.2076.

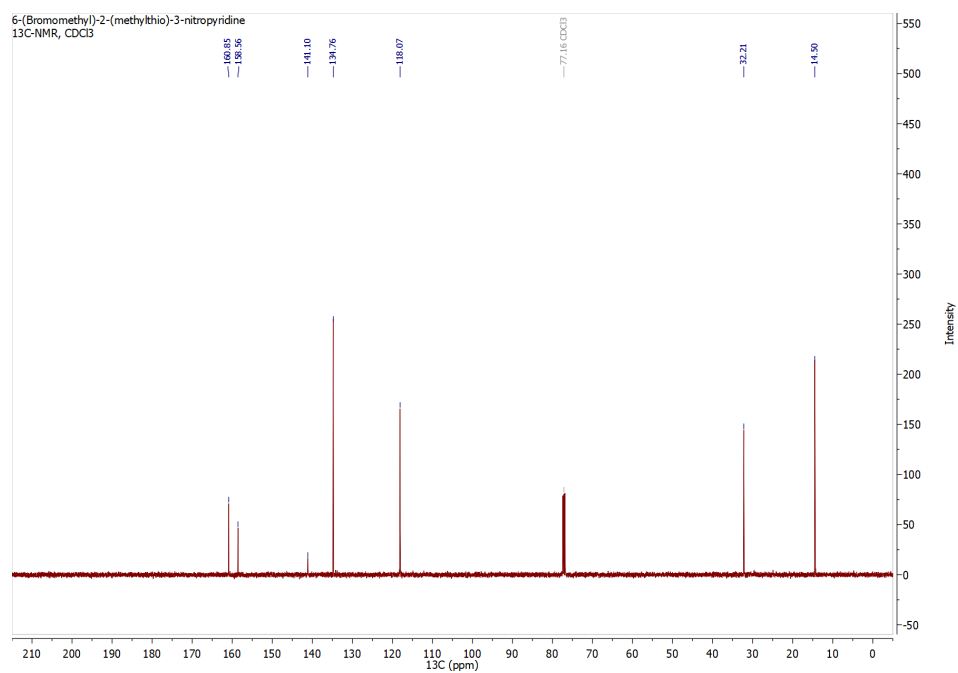
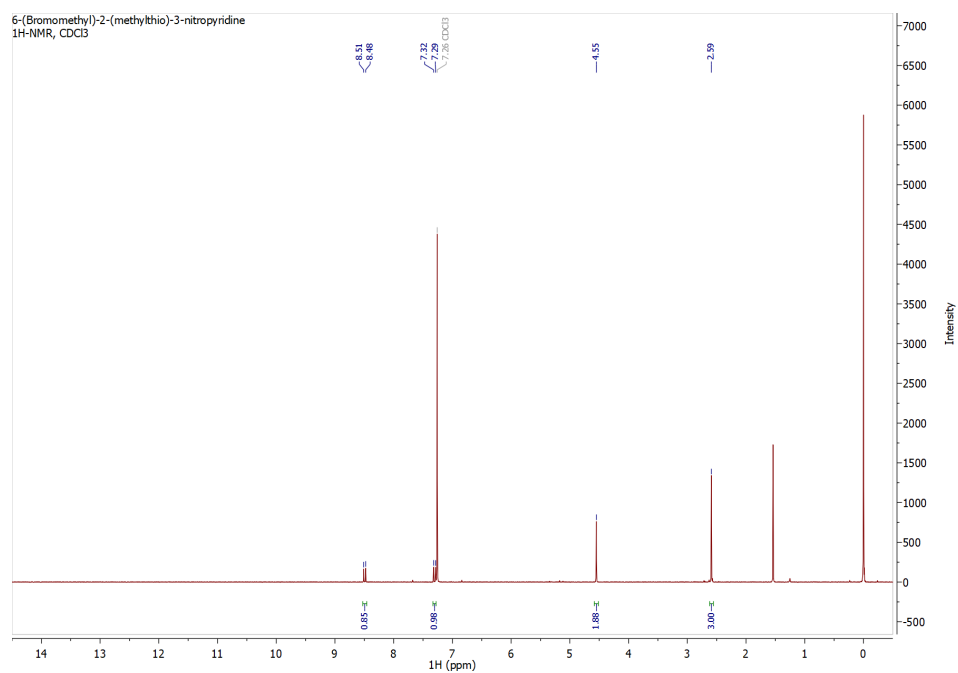
Yb: HRMS: $[M+H]^+$ C₂₈H₄₄N₆O₁₀SYb, m/z (calc.) = 830.2227, m/z (meas.) = 830.2224.

NMR spectra of synthesized compounds

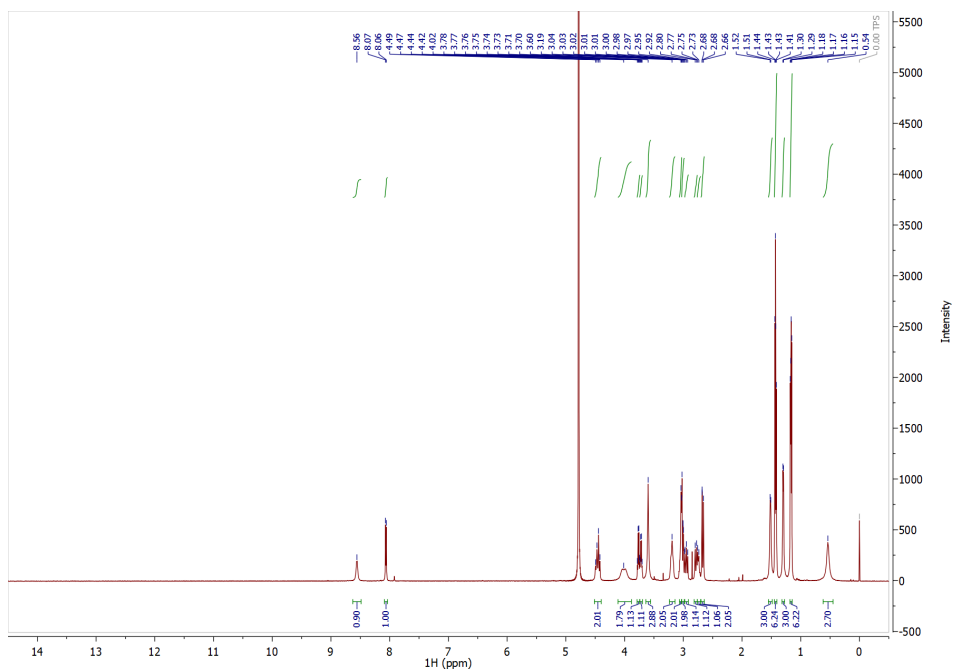
¹H- and ¹³C-NMR of 6-Methyl-2-(methylthio)-3-nitropyridine



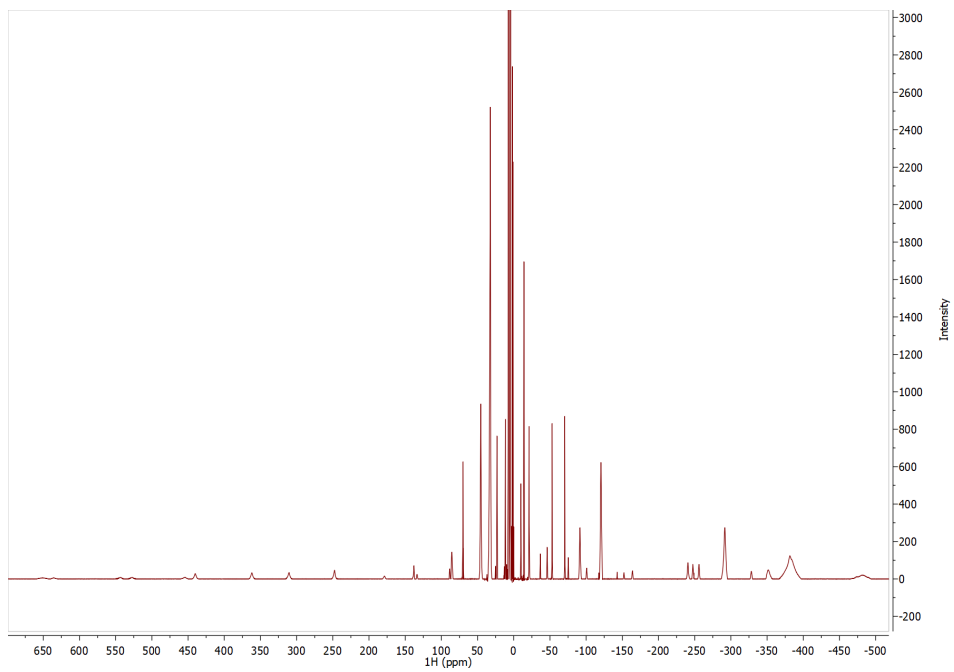
¹H- and ¹³C-NMR of 6-(Bromomethyl)-2-(methylthio)-3-nitropyridine



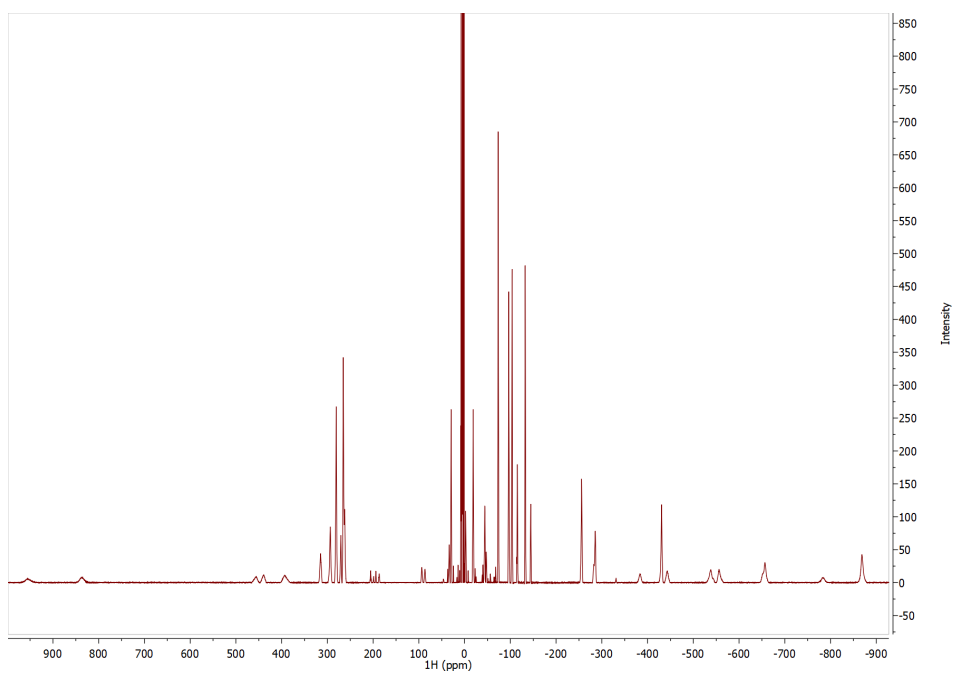
¹H NMR of Lu-M7-Nitro



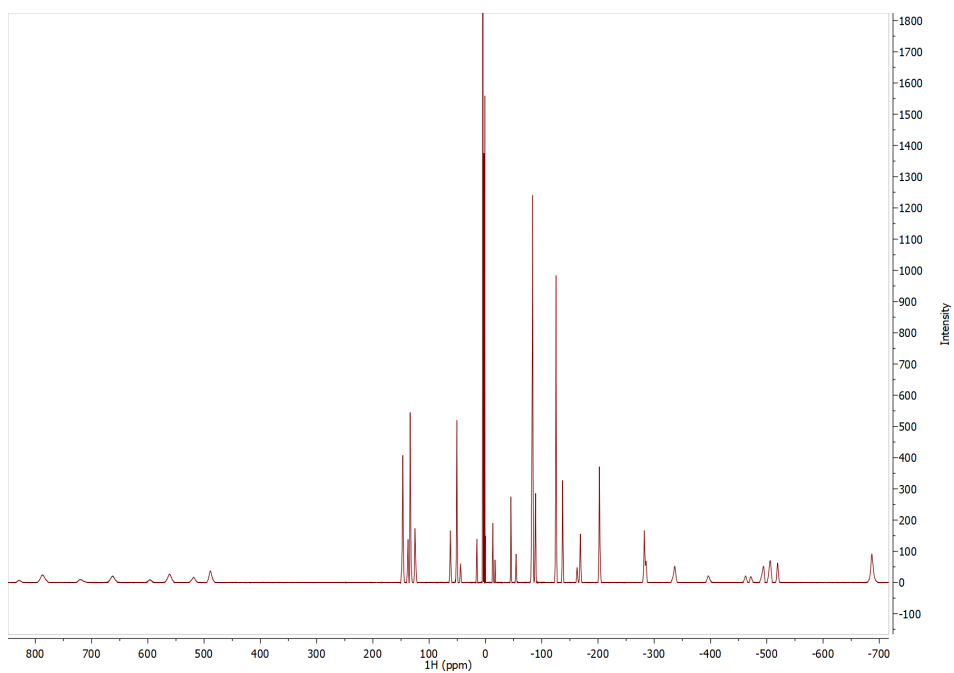
¹H-NMR of Tm-M7-Nitro



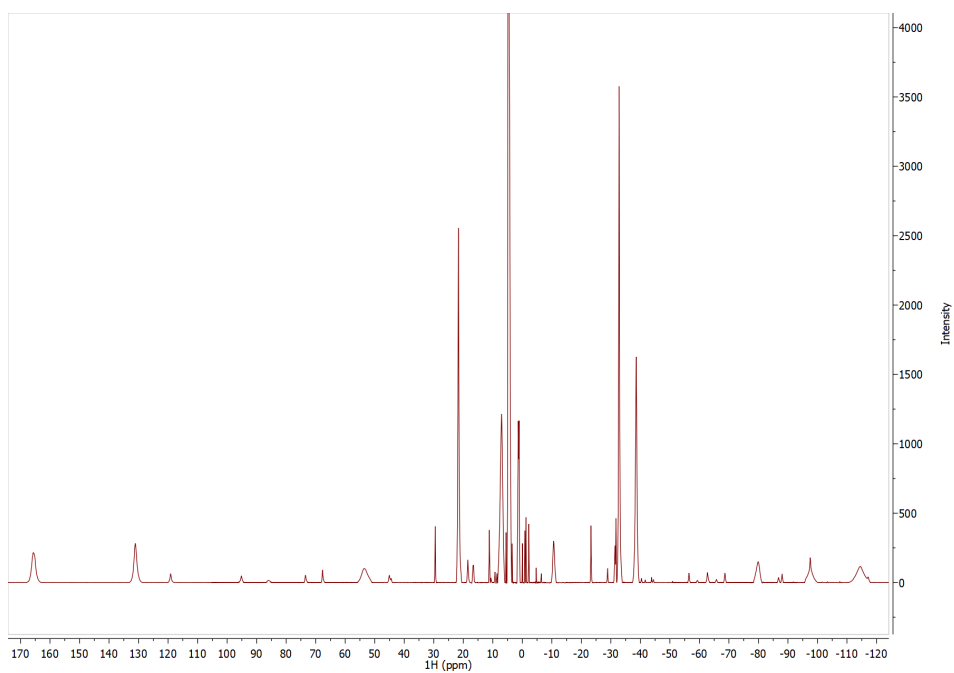
¹H-NMR of Dy-M7-Nitro



¹H-NMR of Tb-M7-Nitro



¹H-NMR of Yb-M7-Nitro



Analytical HPLC-ESI-MS measurement

Lu-M7-Nitro

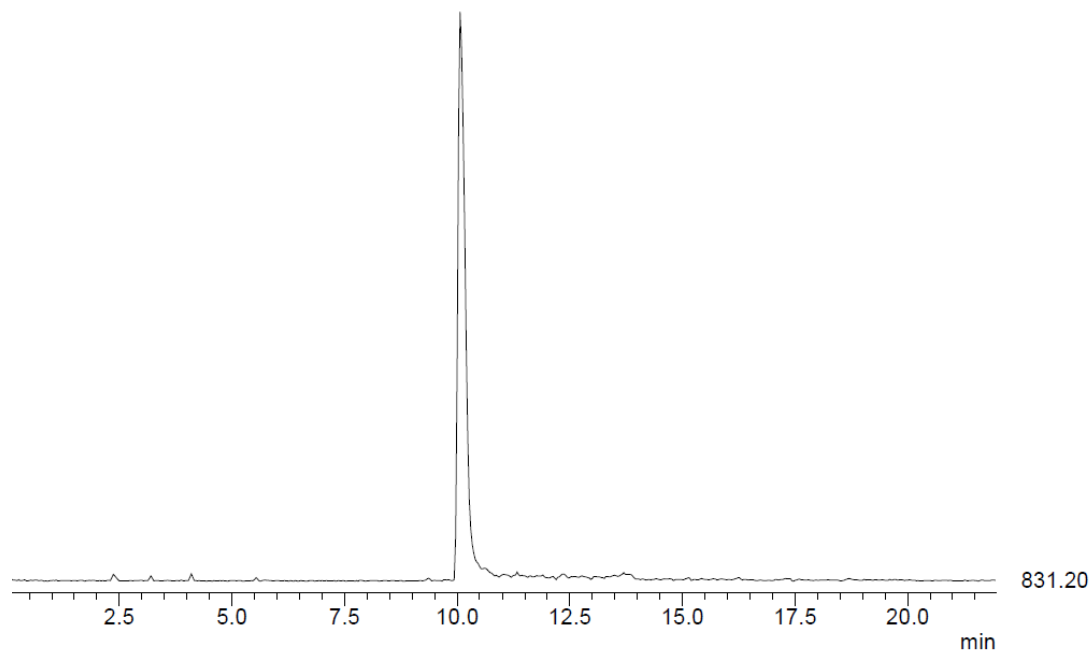


Figure S1: HPLC-ESI-MS trace of Lu-M7-Nitro.

Tm-M7-Nitro

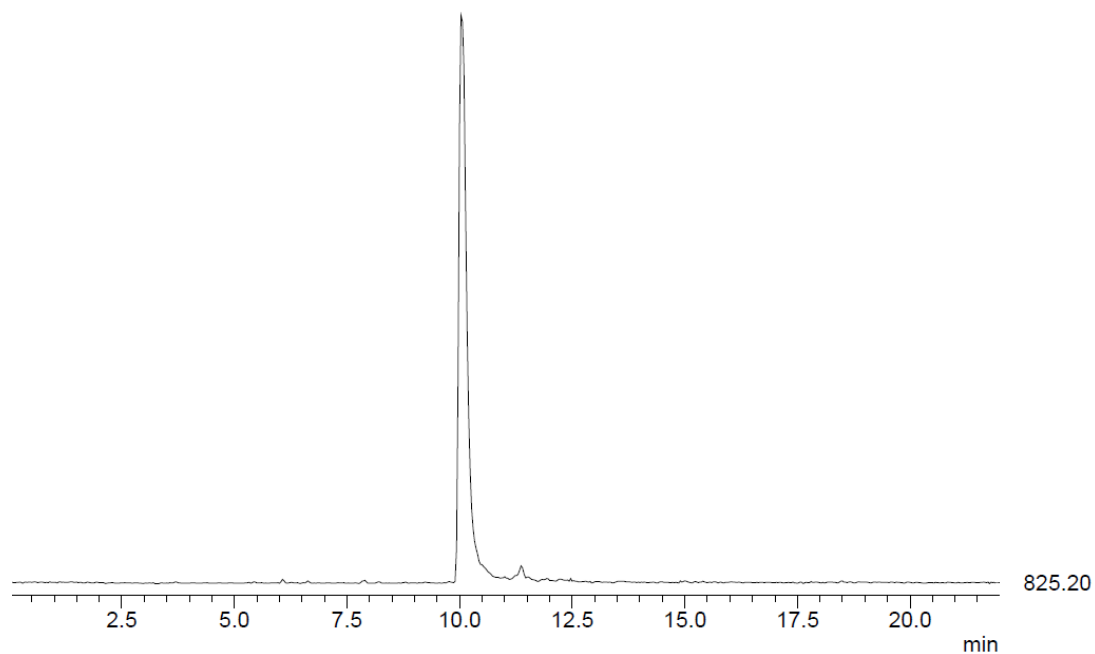


Figure S2: HPLC-ESI-MS trace of Tm-M7-Nitro.

Dy-M7-Nitro

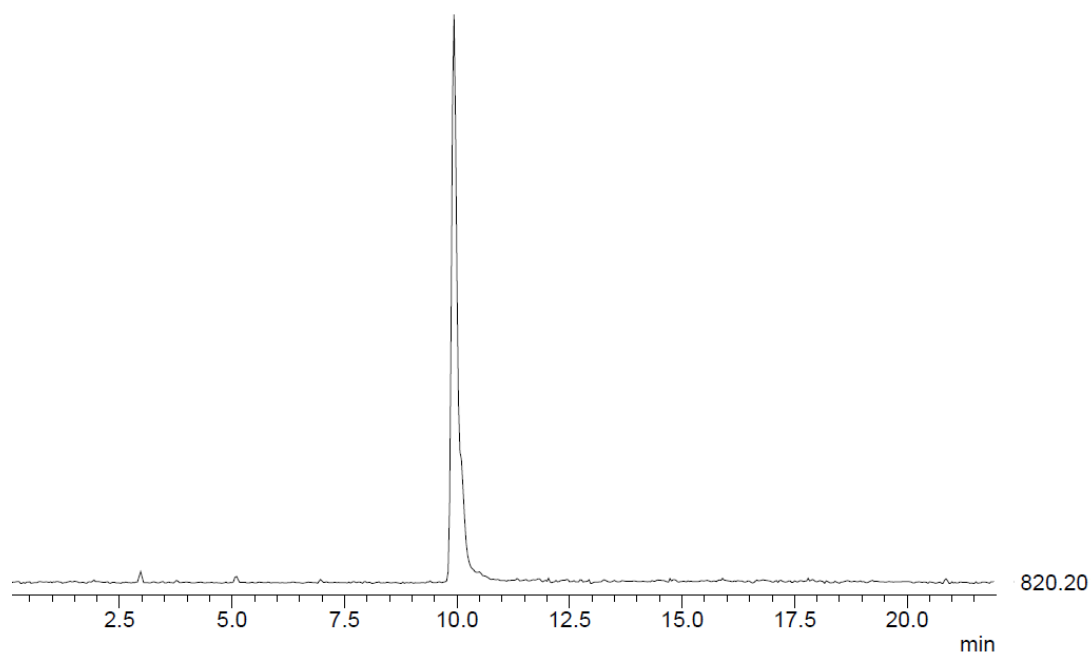


Figure S3: HPLC-ESI-MS trace of Dy-M7-Nitro.

Tb-M7-Nitro

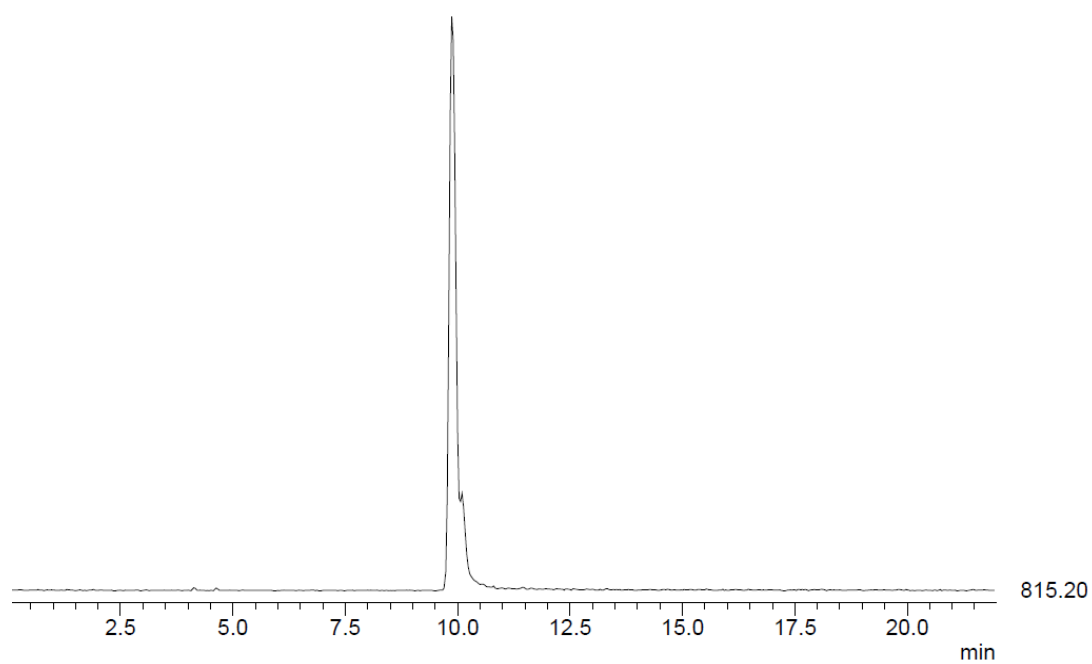


Figure S4: HPLC-ESI-MS trace of Tb-M7-Nitro.

Yb-M7-Nitro

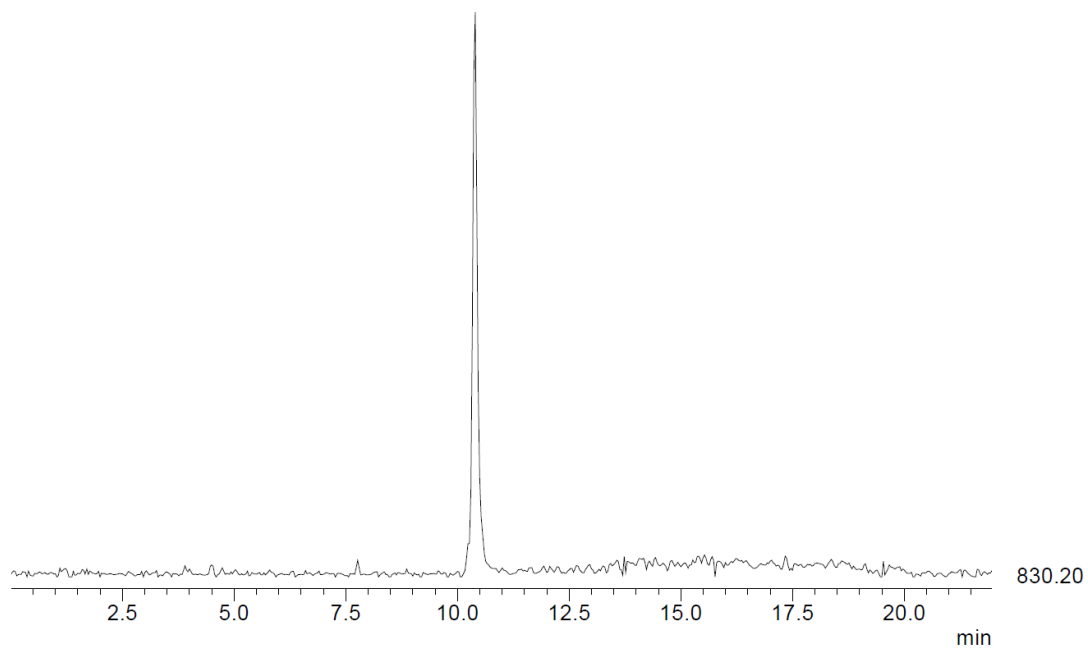


Figure S5: HPLC-ESI-MS trace of Yb-M7-Nitro.

Gd-M7-Nitro

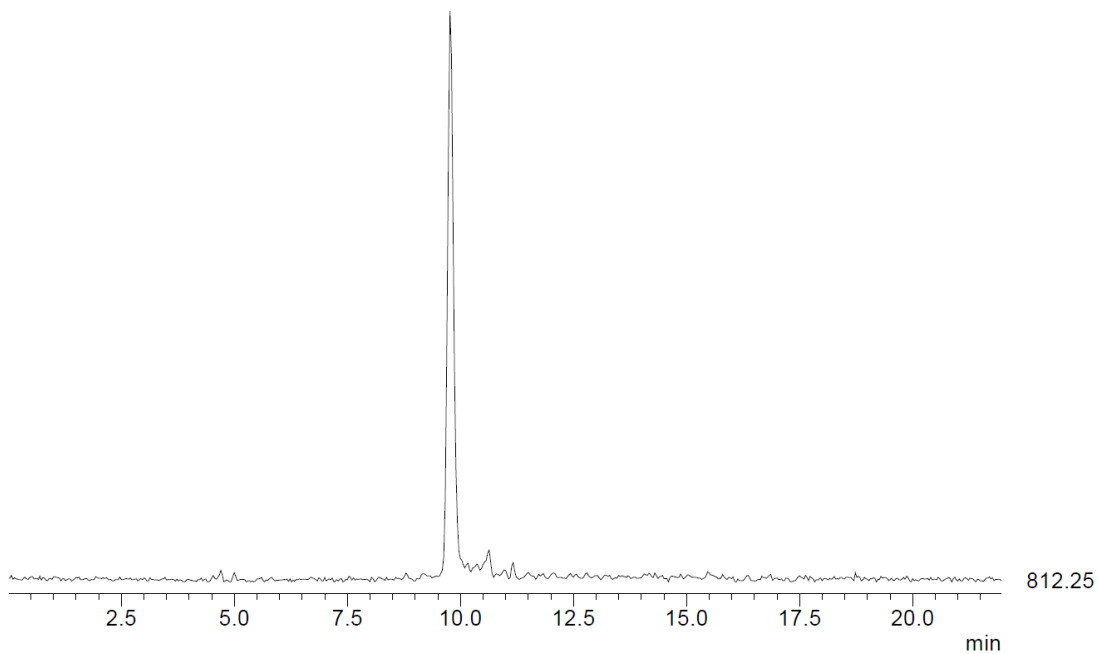


Figure S6: HPLC-ESI-MS trace of Gd-M7-Nitro.

Overlay of ^1H - ^{15}N HSQC spectra of Tm- (red), Dy- (blue), Tb- (orange) and Lu-M7-Nitro-Ub^{S57C} (black)

Measured in 10 mM phosphate buffer with pH 6.0 at a temperature of 298 K on a 600 MHz Bruker Avance III NMR spectrometer equipped with a cryogenic QCI-F probe.

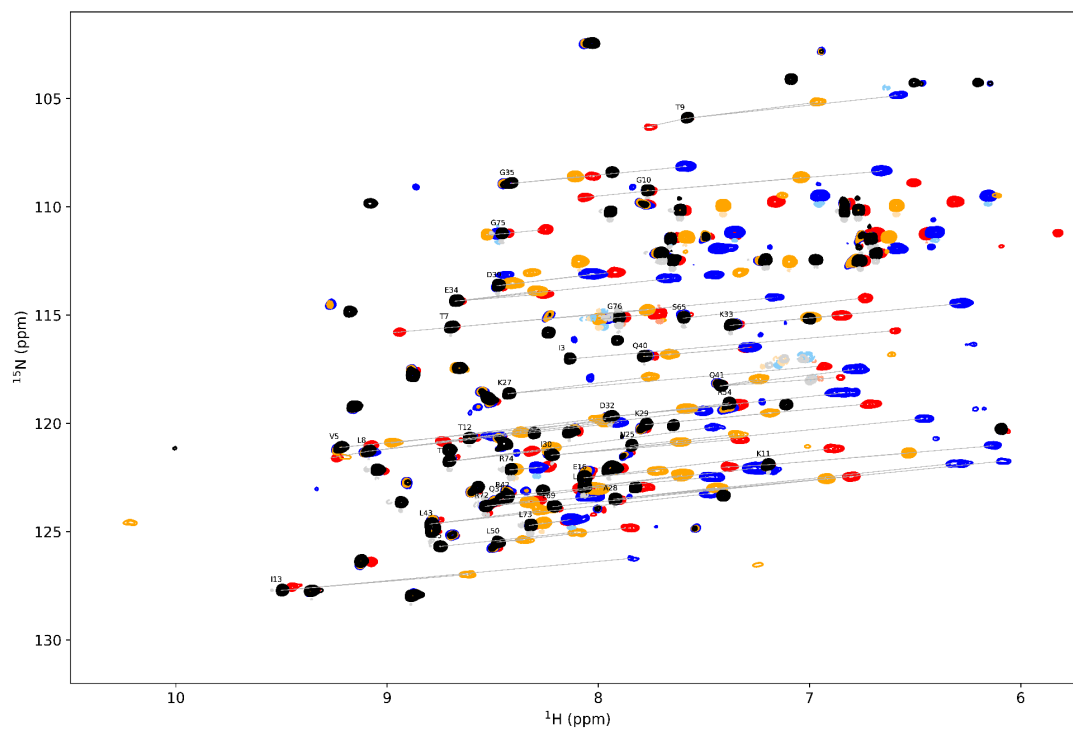


Figure S7: Overlay of ^1H - ^{15}N HSQC spectra of Tm- (red, number of scans = 40), Dy- (blue, number of scans = 100), Tb- (orange, number of scans = 80) and Lu-M7-Nitro-Ub^{S57C} (black, number of scans = 24).

Overlay of ^1H - ^{15}N HSQC spectra of Tm- (red), Dy- (blue), Tb- (orange), Yb- (green) and Lu-M7-Nitro (black) attached to selectively ^{15}N leucine labelled hCA II S50C

Measured in 10 mM phosphate buffer with pH 6.8 at a temperature of 298 K on a 600 MHz Bruker Avance III NMR spectrometer equipped with a cryogenic QCI-F probe.

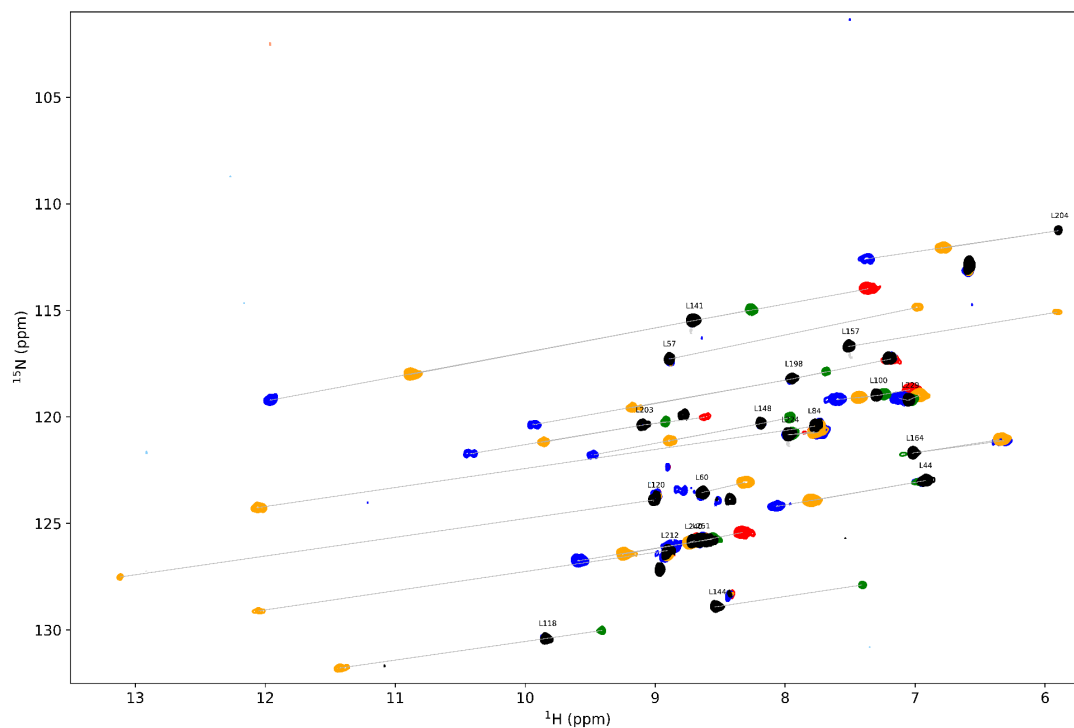


Figure S8: Overlay of ^1H - ^{15}N HSQC spectra of Tm- (red, number of scans = 240), Dy- (blue, number of scans = 288), Tb- (orange, number of scans = 160), Yb- (green, number of scans = 136) and Lu-M7-Nitro (black, number of scans = 72) attached to selectively ^{15}N leucine labelled hCA II S50C.

Shift list comparison of ^1H - ^{15}N HSQC spectra of Tm- and Lu-M7-Nitro-Ub^{S57C}

Table S3: Shift list comparison of ^1H - ^{15}N HSQC spectra of Tm- and Lu-M7-Nitro-Ub^{S57C}.

Reson. 1	Residue	Shift Tm	Shift Lu	ΔPCS	RACS	Reson. 2	Residue	Shift Tm	Shift Lu	ΔPCS	RACS
H	3Ile	6.62	8.14	-1.52	0.002	N	3Ile	115.74	117.03	-1.29	0.017
H	5Val	9.10	9.22	-0.12	0.008	N	5Val	120.99	121.11	-0.12	-0.073
H	7Thr	8.97	8.70	0.27	0.008	N	7Thr	115.80	115.57	0.23	-0.099
H	12Thr	8.77	8.61	0.15	0.003	N	12Thr	120.85	120.69	0.16	-0.006
H	14Thr	8.35	8.71	-0.36	0.003	N	14Thr	121.30	121.77	-0.48	-0.027
H	16Glu	6.92	8.07	-1.15	-0.001	N	16Glu	121.19	122.52	-1.33	-0.081
H	27Lys	6.96	8.43	-1.46	0.003	N	27Lys	117.37	118.63	-1.26	0.024
H	28Ala	6.84	7.93	-1.09	-0.003	N	28Ala	122.47	123.50	-1.03	-0.058
H	29Lys	6.75	7.78	-1.03	0.007	N	29Lys	119.13	120.06	-0.93	-0.046
H	30Ile	7.36	8.22	-0.86	-0.002	N	30Ile	120.75	121.47	-0.72	0.002
H	32Asp	7.38	7.94	-0.57	0.001	N	32Asp	119.16	119.69	-0.53	-0.061
H	33Lys	6.88	7.38	-0.50	0.003	N	33Lys	115.03	115.48	-0.44	-0.064
H	34Glu	8.29	8.68	-0.39	-0.002	N	34Glu	114.05	114.36	-0.30	0.070
H	39Asp	7.95	8.48	-0.53	0.005	N	39Asp	113.05	113.65	-0.60	-0.002
H	40Gln	7.33	7.79	-0.46	-0.003	N	40Gln	116.54	116.92	-0.38	0.059
H	41Gln	6.89	7.42	-0.54	-0.006	N	41Gln	117.90	118.28	-0.38	0.050

Shift list comparison of ^1H - ^{15}N HSQC spectra of Dy- and Lu-M7-Nitro-Ub^{S57C}

Table S4: Shift list comparison of ^1H - ^{15}N HSQC spectra of Dy- and Lu-M7-Nitro-Ub^{S57C}.

Reson. 1	Residue	Shift Dy	Shift Lu	ΔPCS	RACS	Reson. 2	Residue	Shift Dy	Shift Lu	ΔPCS	RACS
H	7Thr	7.18	8.70	-1.52	-0.013	N	7Thr	114.21	115.57	-1.37	0.251
H	12Thr	7.51	8.61	-1.11	-0.007	N	12Thr	119.43	120.69	-1.26	-0.010
H	13Ile	7.85	9.50	-1.65	-0.005	N	13Ile	126.24	127.71	-1.47	0.099
H	14Thr	7.46	8.71	-1.25	0.003	N	14Thr	120.18	121.77	-1.59	-0.124
H	28Ala	6.32	7.93	-1.60	0.013	N	28Ala	121.86	123.50	-1.64	-0.094
H	30Ile	6.47	8.22	-1.75	0.004	N	30Ile	119.79	121.47	-1.68	-0.072
H	32Asp	6.85	7.94	-1.10	0.005	N	32Asp	118.60	119.69	-1.08	-0.061
H	33Lys	6.28	7.38	-1.09	0.002	N	33Lys	114.47	115.48	-1.01	0.031
H	34Glu	7.67	8.68	-1.00	-0.000	N	34Glu	113.33	114.36	-1.03	-0.113
H	39Asp	8.03	8.48	-0.45	-0.002	N	39Asp	113.12	113.65	-0.53	-0.068
H	40Gln	7.28	7.79	-0.50	-0.000	N	40Gln	116.50	116.92	-0.43	-0.034
H	41Gln	6.79	7.42	-0.64	0.007	N	41Gln	117.52	118.28	-0.76	-0.068
H	42Arg	7.47	8.44	-0.97	0.006	N	42Arg	122.49	123.34	-0.86	-0.023
H	43Leu	7.47	8.80	-1.33	-0.001	N	43Leu	123.24	124.65	-1.41	0.076
H	69Leu	6.09	8.22	-2.12	-0.004	N	69Leu	121.75	123.84	-2.09	0.125
H	71Leu	7.27	8.07	-0.80	0.009	N	71Leu	122.08	122.96	-0.88	-0.053

Shift list comparison of ^1H - ^{15}N HSQC spectra of Tb- and Lu-M7-Nitro-Ub^{S57C}

Table S5: Shift list comparison of ^1H - ^{15}N HSQC spectra of Tb- and Lu-M7-Nitro-Ub^{S57C}.

Reson. 1	Residue	Shift Tb	Shift Lu	ΔPCS	RACS	Reson. 2	Residue	Shift Tb	Shift Lu	ΔPCS	RACS
H	5Val	7.98	9.22	-1.25	-0.008	N	5Val	119.91	121.11	-1.21	0.120
H	7Thr	7.77	8.70	-0.93	-0.010	N	7Thr	114.77	115.57	-0.80	0.176
H	12Thr	7.96	8.61	-0.65	-0.005	N	12Thr	119.81	120.69	-0.88	-0.002
H	13Ile	8.61	9.50	-0.89	-0.003	N	13Ile	126.97	127.71	-0.74	0.089
H	14Thr	8.23	8.71	-0.48	-0.000	N	14Thr	121.22	121.77	-0.55	-0.054
H	15Leu	8.11	8.75	-0.64	0.007	N	15Leu	125.03	125.69	-0.66	-0.055
H	16Glu	7.73	8.07	-0.34	0.005	N	16Glu	122.21	122.52	-0.30	0.010
H	25Asn	7.35	7.85	-0.49	-0.002	N	25Asn	120.52	121.02	-0.50	-0.014
H	27Lys	7.76	8.43	-0.66	-0.002	N	27Lys	117.87	118.63	-0.76	-0.084
H	28Ala	7.45	7.93	-0.48	0.008	N	28Ala	122.98	123.50	-0.52	-0.025
H	29Lys	7.19	7.78	-0.59	-0.004	N	29Lys	119.54	120.06	-0.52	0.016
H	30Ile	7.61	8.22	-0.61	0.003	N	30Ile	120.87	121.47	-0.60	-0.040
H	31Gln	8.02	8.47	-0.44	0.003	N	31Gln	123.03	123.53	-0.51	-0.063
H	32Asp	7.59	7.94	-0.36	0.002	N	32Asp	119.33	119.69	-0.36	-0.002
H	33Lys	7.00	7.38	-0.37	-0.001	N	33Lys	115.13	115.48	-0.34	0.049
H	34Glu	8.30	8.68	-0.38	0.001	N	34Glu	113.91	114.36	-0.45	-0.091
H	39Asp	8.41	8.48	-0.07	-0.003	N	39Asp	113.55	113.65	-0.10	-0.025
H	40Gln	7.67	7.79	-0.12	0.000	N	40Gln	116.82	116.92	-0.10	-0.033
H	41Gln	7.25	7.42	-0.18	0.007	N	41Gln	117.97	118.28	-0.31	-0.061
H	43Leu	8.28	8.80	-0.52	0.001	N	43Leu	124.03	124.65	-0.62	0.033
H	69Leu	6.92	8.22	-1.29	-0.003	N	69Leu	122.56	123.84	-1.28	0.067
H	71Leu	7.61	8.07	-0.46	0.008	N	71Leu	122.41	122.96	-0.56	-0.073

Shift list comparison of ^1H - ^{15}N HSQC spectra of Tm- and Lu-M7-Nitro attached to selectively ^{15}N leucine labelled hCA II S50C

Table S6: Shift list comparison of ^1H - ^{15}N HSQC spectra of Tm- and Lu-M7-Nitro attached to selectively ^{15}N leucine labelled hCA II S50C.

Reson. 1	Residue	Shift Tm	Shift Lu	ΔPCS	RACS	Reson. 2	Residue	Shift Tm	Shift Lu	ΔPCS	RACS
H	100Leu	7.04	7.30	-0.26	-0.002	N	100Leu	118.73	119.01	-0.28	-0.092
H	141Leu	7.37	8.72	-1.35	0.004	N	141Leu	114.00	115.50	-1.50	0.006
H	198Leu	7.18	7.95	-0.77	0.000	N	198Leu	117.28	118.23	-0.94	-0.076
H	203Leu	8.63	9.10	-0.47	-0.002	N	203Leu	120.03	120.39	-0.36	0.077
H	251Leu	8.34	8.66	-0.33	0.002	N	251Leu	125.43	125.84	-0.40	-0.097

Shift list comparison of ^1H - ^{15}N HSQC spectra of Dy- and Lu-M7-Nitro attached to selectively ^{15}N leucine labelled hCA II S50C

Table S7: Shift list comparison of ^1H - ^{15}N HSQC spectra of Dy- and Lu-M7-Nitro attached to selectively ^{15}N leucine labelled hCA II S50C.

Reson. 1	Residue	Shift Dy	Shift Lu	ΔPCS	RACS	Reson. 2	Residue	Shift Dy	Shift Lu	ΔPCS	RACS
H	44Leu	8.07	6.92	1.15	0.010	N	44Leu	124.20	122.99	1.22	-0.073
H	100Leu	7.60	7.30	0.30	0.008	N	100Leu	119.19	119.01	0.18	-0.011
H	141Leu	11.96	8.72	3.25	-0.007	N	141Leu	119.22	115.50	3.72	0.171
H	148Leu	9.47	8.19	1.28	-0.014	N	148Leu	121.78	120.32	1.46	0.245
H	164Leu	6.32	7.02	-0.70	-0.004	N	164Leu	121.12	121.70	-0.59	0.073
H	198Leu	9.94	7.95	1.99	0.008	N	198Leu	120.37	118.23	2.15	-0.008
H	203Leu	10.42	9.10	1.32	-0.005	N	203Leu	121.73	120.39	1.34	0.058
H	204Leu	7.37	5.91	1.47	0.004	N	204Leu	112.59	111.27	1.32	-0.036
H	224Leu	7.75	7.98	-0.23	-0.002	N	224Leu	120.74	120.84	-0.11	0.062
H	229Leu	7.12	7.06	0.06	0.011	N	229Leu	119.14	119.22	-0.09	-0.092
H	251Leu	9.59	8.66	0.93	0.009	N	251Leu	126.75	125.84	0.92	-0.060

Shift list comparison of ^1H - ^{15}N HSQC spectra of Tb- and Lu-M7-Nitro attached to selectively ^{15}N leucine labelled hCA II S50C

Table S8: Shift list comparison of ^1H - ^{15}N HSQC spectra of Tb- and Lu-M7-Nitro attached to selectively ^{15}N leucine labelled hCA II S50C.

Reson. 1	Residue	Shift Tb	Shift Lu	ΔPCS	RACS	Reson. 2	Residue	Shift Tb	Shift Lu	ΔPCS	RACS
H	44Leu	7.80	6.92	0.89	0.007	N	44Leu	123.94	122.99	0.95	-0.060
H	57Leu	6.99	8.89	-1.90	-0.002	N	57Leu	114.88	117.29	-2.41	-0.008
H	60Leu	8.31	8.64	-0.33	0.003	N	60Leu	123.08	123.55	-0.48	-0.134
H	84Leu	12.06	7.77	4.28	-0.005	N	84Leu	124.26	120.44	3.82	0.057
H	100Leu	7.44	7.30	0.14	0.003	N	100Leu	119.11	119.01	0.10	0.035
H	118Leu	11.43	9.85	1.58	0.006	N	118Leu	131.79	130.41	1.38	-0.119
H	120Leu	13.12	9.01	4.11	0.009	N	120Leu	127.51	123.90	3.61	-0.097
H	141Leu	10.89	8.72	2.17	-0.006	N	141Leu	118.02	115.50	2.52	0.123
H	148Leu	8.90	8.19	0.71	-0.009	N	148Leu	121.13	120.32	0.81	0.166
H	157Leu	5.91	7.51	-1.60	0.002	N	157Leu	115.09	116.70	-1.61	-0.037
H	164Leu	6.35	7.02	-0.67	-0.001	N	164Leu	121.07	121.70	-0.63	0.029
H	198Leu	9.18	7.95	1.23	0.003	N	198Leu	119.57	118.23	1.35	0.038
H	203Leu	9.85	9.10	0.75	-0.004	N	203Leu	121.18	120.39	0.79	0.050
H	204Leu	6.78	5.91	0.88	0.001	N	204Leu	112.08	111.27	0.81	-0.011
H	212Leu	12.05	8.90	3.15	0.004	N	212Leu	129.11	126.27	2.83	-0.134
H	224Leu	7.78	7.98	-0.20	-0.001	N	224Leu	120.70	120.84	-0.15	0.007
H	229Leu	6.97	7.06	-0.09	0.009	N	229Leu	119.00	119.22	-0.22	-0.095
H	240Leu	8.74	8.72	0.01	0.001	N	240Leu	125.91	125.86	0.05	0.018
H	251Leu	9.25	8.66	0.59	0.006	N	251Leu	126.44	125.84	0.61	-0.025

Shift list comparison of ^1H - ^{15}N HSQC spectra of Yb- and Lu-M7-Nitro attached to selectively ^{15}N leucine labelled hCA II S50C

Table S9: Shift list comparison of ^1H - ^{15}N HSQC spectra of Yb- and Lu-M7-Nitro attached to selectively ^{15}N leucine labelled hCA II S50C.

Reson. 1	Residue	Shift Yb	Shift Lu	ΔPCS	RACS	Reson. 2	Residue	Shift Yb	Shift Lu	ΔPCS	RACS
H	100Leu	7.24	7.30	-0.06	-0.001	N	100Leu	118.96	119.01	-0.05	-0.016
H	118Leu	9.41	9.85	-0.43	-0.001	N	118Leu	130.03	130.41	-0.38	0.020
H	141Leu	8.27	8.72	-0.45	0.001	N	141Leu	114.99	115.50	-0.51	-0.007
H	144Leu	7.41	8.54	-1.13	-0.001	N	144Leu	127.89	128.92	-1.03	0.025
H	148Leu	7.97	8.19	-0.22	0.002	N	148Leu	120.04	120.32	-0.28	-0.034
H	198Leu	7.69	7.95	-0.26	-0.000	N	198Leu	117.89	118.23	-0.34	-0.014
H	203Leu	8.92	9.10	-0.18	-0.000	N	203Leu	120.23	120.39	-0.16	0.012
H	224Leu	7.96	7.98	-0.02	0.001	N	224Leu	120.81	120.84	-0.03	-0.011
H	229Leu	7.04	7.06	-0.01	-0.001	N	229Leu	119.22	119.22	0.00	0.015
H	240Leu	8.70	8.72	-0.03	0.001	N	240Leu	125.82	125.86	-0.04	-0.010
H	251Leu	8.56	8.66	-0.10	-0.001	N	251Leu	125.75	125.84	-0.09	-0.004

Residual dipolar couplings measured in ^1H - ^{15}N HSQC spectra of Tm- and Lu-M7-Nitro-Ub^{S57C}

Table S10: Residual dipolar couplings measured in ^1H - ^{15}N HSQC spectra of Tm- and Lu-M7-Nitro-Ub^{S57C}.

Residue	RDC (Hz)
3	-5.3
5	-18.9
7	-15.0
12	-5.7
14	-11.8
16	-14.8
27	0.3
28	8.1
29	-13.2
30	-4.2
32	1.7
33	-11.4
34	3.0
39	-4.9
40	-8.4
41	26.4
70	25.7

Residual dipolar couplings measured in ^1H - ^{15}N HSQC spectra of Dy- and Lu-M7-Nitro-Ub^{S57C}

Table S11: Residual dipolar couplings measured in ^1H - ^{15}N HSQC spectra of Dy- and Lu-M7-Nitro-Ub^{S57C}.

Residue	RDC (Hz)
7	41.3
12	20.2
13	33.1
14	-16.7
28	-20.0
30	-7.4
32	-30.3
33	-11.3
34	-0.2
39	-25.6
40	18.2
41	-28.7
43	-5.3
69	23.1
71	-25.8

Residual dipolar couplings measured in ^1H - ^{15}N HSQC spectra of Tb- and Lu-M7-Nitro-Ub^{S57C}

Table S12: Residual dipolar couplings measured in ^1H - ^{15}N HSQC spectra of Tb- and Lu-M7-Nitro-Ub^{S57C}.

Residue	RDC (Hz)
5	13.4
7	28.5
12	12.3
13	25.7
14	-2.6
15	-8.9
16	-2.9
27	-2.7
28	-16.7
29	-5.9
30	-6.3
31	-16.4
32	-15.3
33	0.5
34	-8.9
39	-11.2
40	16.0
41	-35.9
43	-11.9
69	8.9
71	-29.1

Residual dipolar couplings measured in ^1H - ^{15}N HSQC spectra of Tb- and Lu-M7-Nitro attached to selectively ^{15}N leucine labelled hCA II S50C

Table S13: Residual dipolar couplings measured in ^1H - ^{15}N HSQC spectra of Tb- and Lu-M7-Nitro attached to selectively ^{15}N leucine labelled hCA II S50C.

Residue	RDC (Hz)
44	-11.9
57	19.4
60	-24.7
84	31.2
100	-5.7
118	-23.4
120	-28.9
141	25.3
148	32.6
157	-19.2
164	-11.5
203	15.4
204	-9.3
212	-23.7
224	12.9
229	-29.5
240	-0.3
251	-16.1

Residual dipolar couplings measured in ^1H - ^{15}N HSQC spectra of Yb- and Lu-M7-Nitro attached to selectively ^{15}N leucine labelled hCA II S50C

Table S14: Residual dipolar couplings measured in ^1H - ^{15}N HSQC spectra of Yb- and Lu-M7-Nitro attached to selectively ^{15}N leucine labelled hCA II S50C.

Residue	RDC (Hz)
100	-0.7
118	4.7
141	-3.7
144	9.9
148	-2.5
198	0.0
203	-2.1
224	-4.1
229	4.9
240	-0.8
251	0.6

Paramagnetic relaxation enhancements measured in ^1H - ^{15}N HSQC spectra of Gd- and Lu-M7-Nitro-Ub^{S57C}

Table S15: Paramagnetic relaxation enhancements measured in ^1H - ^{15}N HSQC spectra of Gd- and Lu-M7-Nitro-Ub^{S57C}.

Residue	Intensity Lu	Intensity Gd	$I_{\text{para}}/I_{\text{dia}}$ normalized to most distant residue (G76)	Distance to metal centre (Å)
3	86917	0	0.000	18.3
4	94962	49663	0.638	17.7
5	118292	63267	0.653	21.5
6	118676	79359	0.816	21.3
7	155439	115477	0.906	25.8
8	113994	84898	0.909	27.7
9	82717	74156	1.094	29.6
10	152503	129880	1.039	29.3
11	166971	123292	0.901	29.3
12	133538	103833	0.949	27.9
13	104644	75178	0.877	24.4
14	123349	99493	0.984	25.0
15	89614	28104	0.383	20.7
16	174689	91380	0.638	21.4
17	109940	57837	0.642	17.4
18	88948	897	0.012	15.8
20	84887	0	0.000	12.3
21	123910	16036	0.158	13.3
23	80998	5762	0.087	15.4
25	99596	11761	0.144	18.7
27	96665	23125	0.292	20.1
28	167410	97737	0.712	22.3
29	125868	70722	0.686	22.5
30	126145	76504	0.740	23.1
32	145289	114151	0.959	26.9
33	145191	111764	0.939	27.3
34	139282	106794	0.936	28.4
35	84185	68028	0.986	29.8
36	98417	76456	0.948	29.3
39	218771	169435	0.945	28.1
40	126676	105357	1.015	28.2
41	103959	80099	0.940	26.2
44	135114	60234	0.544	18.9
47	103235	50773	0.600	16.4
49	135447	10434	0.094	17.2
50	87124	7731	0.108	16.9
51	76591	2863	0.046	15.0
52	120904	54823	0.553	18.7
54	92930	6823	0.090	15.2

55	47919	1214	0.031	11.2
56	91153	9201	0.123	11.5
57	69446	0	0.000	9.0
58	71597	1143	0.019	8.0
59	73233	0	0.000	7.7
60	123061	0	0.000	5.6
61	78265	0	0.000	7.4
62	74016	0	0.000	9.6
64	76036	1319	0.021	15.3
65	112207	16038	0.174	14.3
66	107603	49296	0.559	15.7
67	118374	46029	0.474	17.8
68	120869	50009	0.505	18.4
69	137858	84595	0.749	22.6
70	114833	67884	0.721	23.3
71	193450	147342	0.929	27.8
72	205409	142694	0.848	28.4
73	274191	206123	0.917	31.9
74	242431	199996	1.007	34.1
75	150933	132985	1.075	36.1
76	461149	377931	1.000	39.4

Paramagnetic relaxation enhancements measured in ^1H - ^{15}N HSQC spectra of Gd- and Lu-M7-Nitro attached to ^{15}N labelled hCA II S50C

Table S16: Paramagnetic relaxation enhancements measured in ^1H - ^{15}N HSQC spectra of Gd- and Lu-M7-Nitro attached to ^{15}N labelled hCA II S50C.

Residue	Intensity Lu	Intensity Gd	$I_{\text{para}}/I_{\text{dia}}$ normalized to most distant residue (E238)	Distance to metal centre (Å)
22	74623	75321	1.051	39.5
23	62717	68511	1.138	38.8
25	61088	65436	1.116	38.0
26	72277	69445	1.001	37.4
27	79305	80715	1.060	35.1
28	58286	62625	1.119	34.2
29	23707	22619	0.994	31.4
31	33405	31774	0.991	29.1
33	46672	43983	0.981	29.8
34	68271	57928	0.884	31.3
35	51704	36232	0.730	31.1
37	88764	82797	0.971	32.9
40	89276	55918	0.652	28.7
41	40502	33931	0.873	25.1
43	95228	63021	0.689	25.5
54	60620	1819	0.031	8.9
56	41231	9283	0.234	14.7
57	38824	14620	0.392	18.9
59	50088	29173	0.607	24.1
62	23533	25004	1.107	29.3
66	26642	19352	0.756	26.3
67	30026	20875	0.724	25.3
69	54706	26419	0.503	20.7
70	45351	17564	0.403	19.0
71	42012	19688	0.488	17.3
72	47711	16093	0.351	19.6
74	77505	30089	0.404	18.0
75	27179	7556	0.290	16.5
81	62800	0	0.000	15.6
82	24552	4258	0.181	16.0
85	30999	11160	0.375	21.3
86	158273	54966	0.362	19.6
92	41724	16018	0.400	18.6
93	33642	16343	0.506	18.7
94	43742	20857	0.497	20.3
97	54850	53477	1.015	28.9
98	21431	17059	0.829	30.0
99	30406	30066	1.030	34.5

100	70619	71283	1.051	34.3
101	57928	61177	1.100	34.0
102	84379	80530	0.994	34.2
103	98626	94106	0.994	32.7
104	34876	34020	1.016	31.4
105	71130	63085	0.924	29.1
106	20623	19947	1.007	27.0
107	25101	23973	0.995	26.8
108	54494	49089	0.938	29.3
109	63981	51017	0.830	29.6
110	61568	53578	0.906	30.9
111	54948	48030	0.910	32.6
112	104955	105335	1.045	32.1
113	69674	57923	0.866	32.4
114	39693	36525	0.958	28.5
115	53452	44706	0.871	27.4
116	82317	65450	0.828	26.1
118	38099	19434	0.531	20.9
119	42069	21259	0.526	20.6
121	36114	17844	0.515	18.1
123	41133	16004	0.405	18.8
124	30000	13708	0.476	22.0
125	62107	24729	0.415	19.9
127	123116	55485	0.469	22.3
128	91008	55895	0.640	24.2
129	67718	42235	0.650	24.4
130	114626	79137	0.719	26.2
131	56161	31646	0.587	24.7
132	91109	64876	0.742	27.4
133	93306	68773	0.768	28.5
134	89308	69136	0.806	26.9
135	88403	71395	0.841	27.7
136	92563	73933	0.832	29.8
137	99502	83720	0.876	29.3
139	81587	63235	0.807	27.9
140	48472	41454	0.891	26.4
144	32326	1656	0.053	17.5
145	43193	18553	0.447	18.8
146	38143	22434	0.613	19.5
147	31606	16539	0.545	20.4
148	43719	24553	0.585	23.2
149	54227	29962	0.575	23.0
151	79568	50122	0.656	25.8
152	40412	27392	0.706	24.7

157	72866	43894	0.627	17.6
158	80638	36824	0.476	19.0
159	102909	43388	0.439	19.1
160	78076	36683	0.489	19.6
161	59362	28099	0.493	22.0
162	92800	53683	0.602	23.6
163	80712	38776	0.500	24.4
164	62629	41402	0.688	26.3
166	98815	81331	0.857	29.5
167	87995	81736	0.967	30.0
168	80262	68337	0.887	32.6
170	42870	41936	1.019	35.6
172	105550	89609	0.884	33.0
173	101695	88075	0.902	31.1
174	71073	52274	0.766	26.8
175	158562	113643	0.746	26.7
177	106445	43475	0.425	20.7
178	79473	27453	0.360	16.5
183	46980	0	0.000	10.4
184	53975	4085	0.079	11.8
185	81778	2353	0.030	11.5
188	50972	5327	0.109	14.9
189	66817	0	0.000	13.5
192	44522	22448	0.525	21.1
196	55783	42014	0.784	23.8
198	33247	28416	0.890	29.0
200	51206	53710	1.092	29.6
203	28423	27634	1.013	32.2
204	20166	16454	0.850	31.6
205	53457	44516	0.867	32.5
206	55938	43709	0.814	28.9
207	63655	55784	0.913	27.4
208	47268	25937	0.571	23.3
209	48486	32593	0.700	24.0
210	41455	19195	0.482	20.3
211	58309	31747	0.567	20.8
212	45310	23590	0.542	18.7
214	52087	28507	0.570	18.6
216	51491	18265	0.369	19.5
217	72034	28788	0.416	19.6
218	68393	38391	0.585	23.2
219	59106	38853	0.685	24.0
220	96015	70280	0.762	28.4
221	103591	78335	0.788	28.8

222	66193	51154	0.805	26.8
223	72891	49840	0.712	27.1
224	100679	78675	0.814	29.5
225	98132	81529	0.865	29.1
226	83982	63901	0.792	28.1
227	55840	46757	0.872	30.5
228	79371	80946	1.062	31.8
229	79067	77730	1.024	31.8
230	53307	46444	0.907	33.5
232	57520	56777	1.028	37.8
233	55337	56855	1.070	39.7
235	34026	34893	1.068	42.1
236	108892	111167	1.063	42.6
238	99806	95831	1.000	43.8
239	57104	58780	1.072	39.9
240	86884	76780	0.920	39.2
242	62157	64547	1.082	34.9
244	59564	57439	1.004	31.9
245	69199	67054	1.009	32.7
248	77481	68119	0.916	34.0
251	87598	88934	1.057	36.4
252	104976	102303	1.015	37.5
254	121797	115014	0.983	35.3
255	101304	83352	0.857	31.8
256	89030	77306	0.904	30.6
257	58471	45890	0.817	26.3
258	80508	49231	0.637	25.9
259	50401	18495	0.382	21.7
260	77598	42160	0.566	22.6
261	120162	70776	0.613	23.6

Residue-specific plots of experimental PRE onto the sequence of ubiquitin S57C and hCA S50C constructs

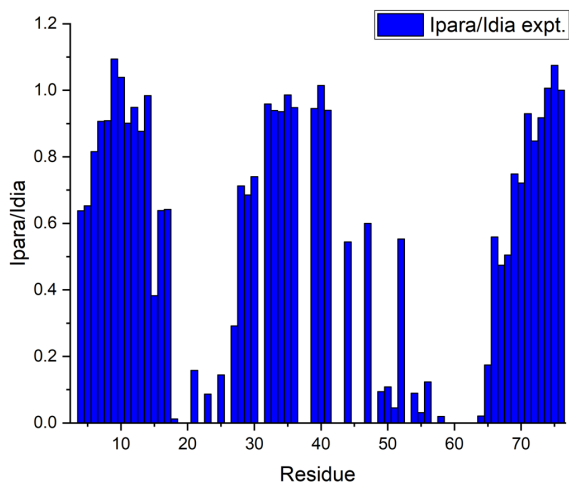


Figure S9: Residue-specific plot of experimental PRE detected on ¹⁵N labelled ubiquitin S57C labelled with Gd-M7-Nitro.

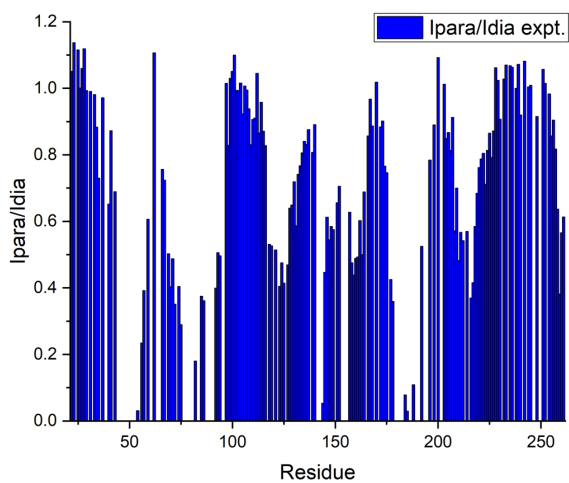


Figure S10: Residue-specific plot of experimental PRE detected on selectively ¹⁵N labelled hCA S50C labelled with Gd-M7-Nitro.

Plots of experimental and calculated PRE for ubiquitin S57C and hCA S50C constructs

Theoretical PRE were calculated using the relation $I_{para}/I_{dia}, \frac{I_{para}}{I_{dia}} = \frac{R_2^{dia} \cdot \exp(-R_2^{para} \cdot t)}{R_2^{dia} + R_2^{para}}$ reported in literature.^{1, 11}

While R_2^{dia} was determined from the linewidths of the diamagnetic sample ($R_2^{dia} = \pi \cdot \Delta\nu_{1/2}$), t is the INEPT delay (9 ms).

R_2^{para} was substituted with the expression from the simplified Solomon-Bloembergen equation:¹¹

$$R_2^{para} = \frac{1}{15} \left(\frac{\mu_0}{4\pi} \right)^2 \frac{\gamma_H^2 g^2 \mu_B^2 J(J+1)}{r^6} \left(4\tau_c + \frac{3\tau_c}{1 + \omega_H^2 \tau_c^2} \right)$$

yielding the complete function for calculation of I_{para}/I_{dia} for a given r :

$$\frac{I_{para}}{I_{dia}} = \frac{R_2^{dia} \cdot \exp\left(-\frac{1}{15} \left(\frac{\mu_0}{4\pi} \right)^2 \frac{\gamma_H^2 g^2 \mu_B^2 J(J+1)}{r^6} \left(4\tau_c + \frac{3\tau_c}{1 + \omega_H^2 \tau_c^2} \right) \cdot t\right)}{R_2^{dia} + \frac{1}{15} \left(\frac{\mu_0}{4\pi} \right)^2 \frac{\gamma_H^2 g^2 \mu_B^2 J(J+1)}{r^6} \left(4\tau_c + \frac{3\tau_c}{1 + \omega_H^2 \tau_c^2} \right)}$$

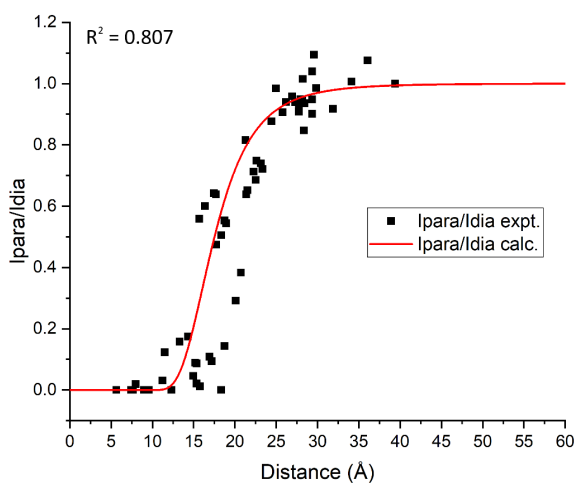


Figure S11: Plot of experimental and calculated PRE of ¹⁵N labelled ubiquitin S57C labelled with Gd-M7-Nitro.

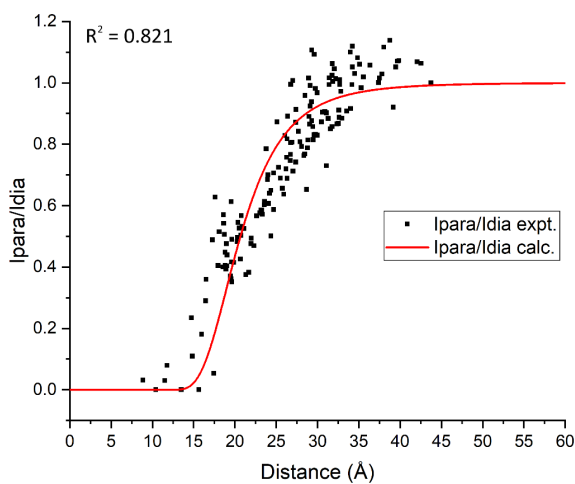


Figure S12: Plot of experimental and calculated PRE of selectively ¹⁵N labelled hCA S50C labelled with Gd-M7-Nitro.

Correlation plots of experimental and back-calculated PCS and RDC

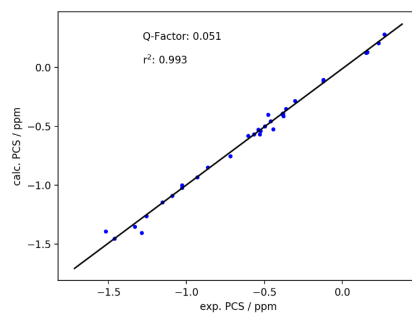


Figure S13: PCS correlation plot of ¹⁵N labelled ubiquitin S57C labelled with Tm-M7-Nitro.

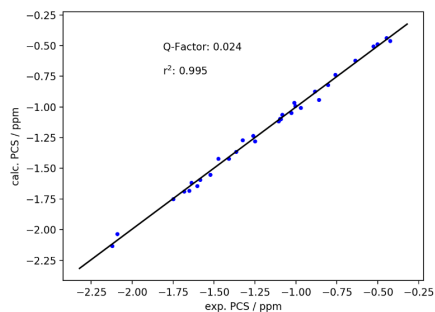


Figure S14: PCS correlation plot of ¹⁵N labelled ubiquitin S57C labelled with Dy-M7-Nitro.

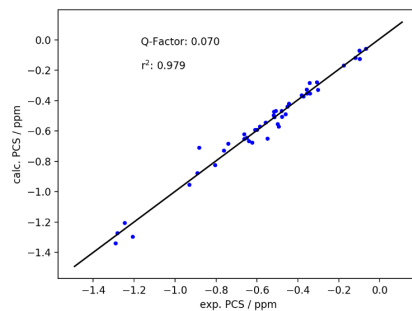


Figure S15: PCS correlation plot of ¹⁵N labelled ubiquitin S57C labelled with Tb-M7-Nitro.

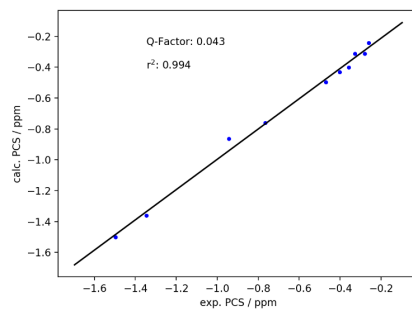


Figure S16: PCS correlation plot of selectively ¹⁵N leucine labelled hCA S50C labelled with Tm-M7-Nitro.

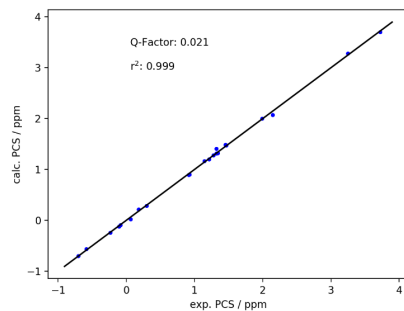


Figure S17: PCS correlation plot of selectively ^{15}N leucine labelled hCA S50C labelled with Dy-M7-Nitro.

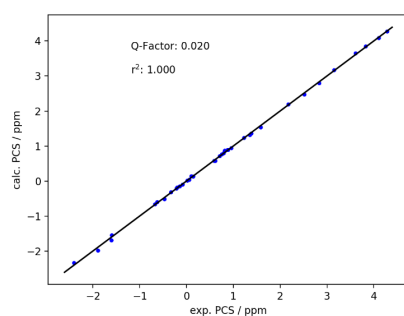


Figure S18: PCS correlation plot of selectively ^{15}N leucine labelled hCA S50C labelled with Tb-M7-Nitro.

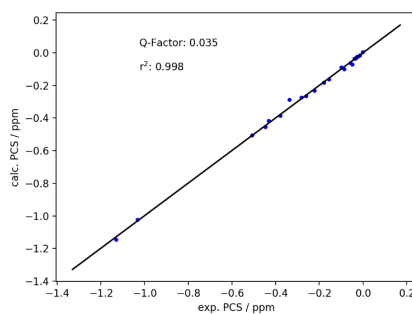


Figure S19: PCS correlation plot of selectively ^{15}N leucine labelled hCA S50C labelled with Yb-M7-Nitro.

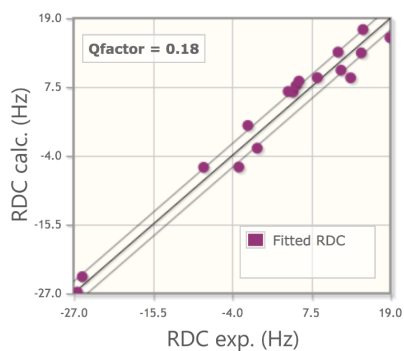


Figure S20: RDC correlation plot of ^{15}N labelled ubiquitin S57C labelled with Tm-M7-Nitro (graphics adapted from Fanten⁹).

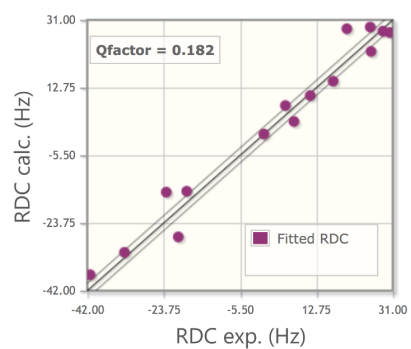


Figure S21: RDC correlation plot of ^{15}N labelled ubiquitin S57C labelled with Dy-M7-Nitro (graphics adapted from Fanten⁹).

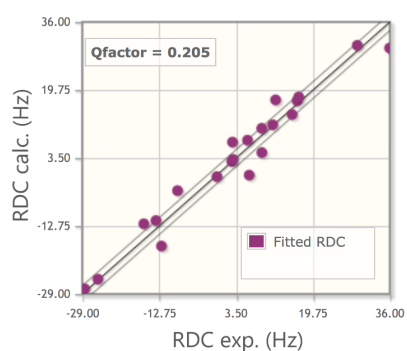


Figure S22: RDC correlation plot of ^{15}N labelled ubiquitin S57C labelled with Tb-M7-Nitro (graphics adapted from Fanten⁹).

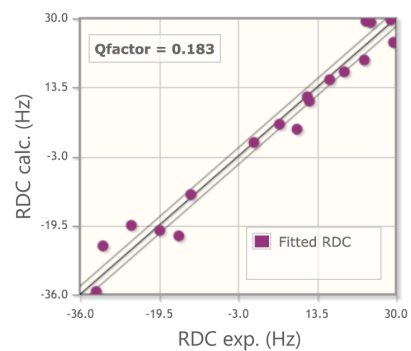


Figure S23: RDC correlation plot of selectively ^{15}N leucine labelled hCA S50C labelled with Tb-M7-Nitro (graphics adapted from Fanten⁹).

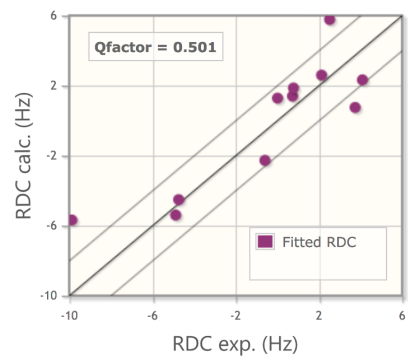


Figure S24: RDC correlation plot of selectively ^{15}N leucine labelled hCA S50C labelled with Yb-M7-Nitro (graphics adapted from Fanten⁹).

ESI-MS measurements of tagging reactions

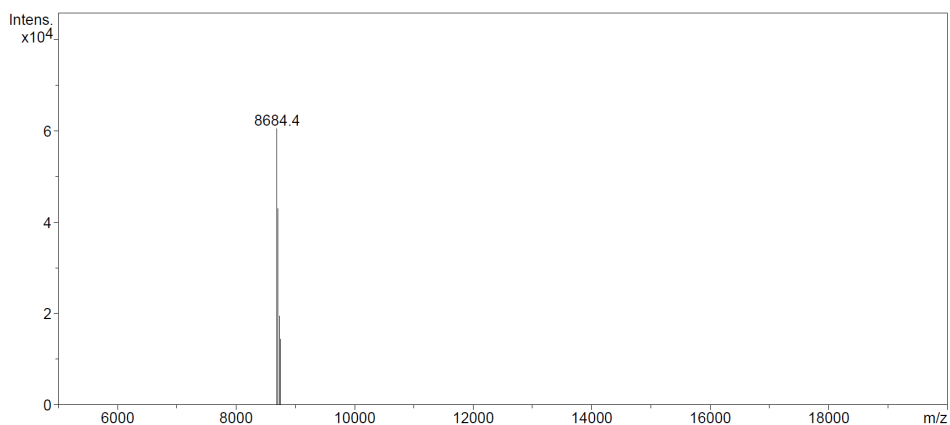


Figure S25: ESI-MS spectrum of ¹⁵N labelled ubiquitin S57C.

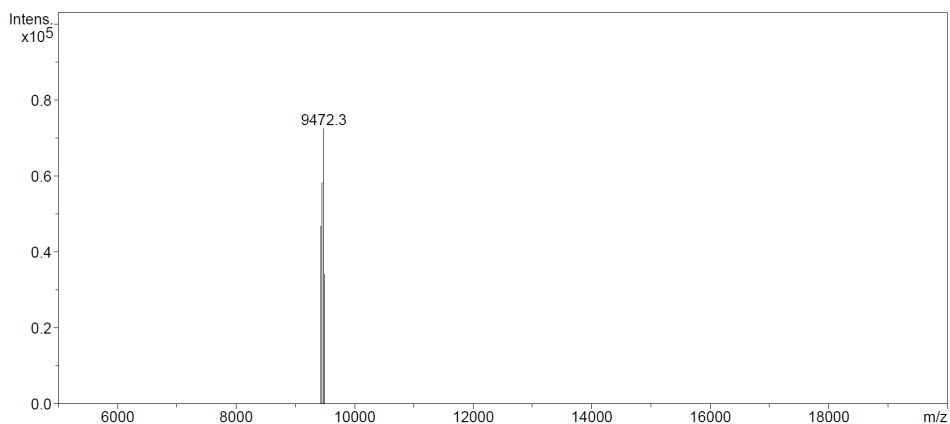


Figure S26: Confirmation of tagging reaction of ¹⁵N labelled ubiquitin S57C with Tm-M7-Nitro monitored by ESI-MS.

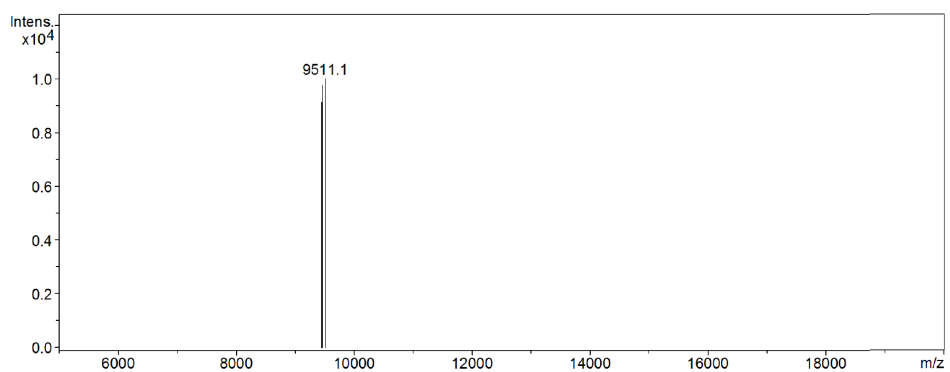


Figure S27: Confirmation of tagging reaction of ¹⁵N labelled ubiquitin S57C with Dy-M7-Nitro monitored by ESI-MS.

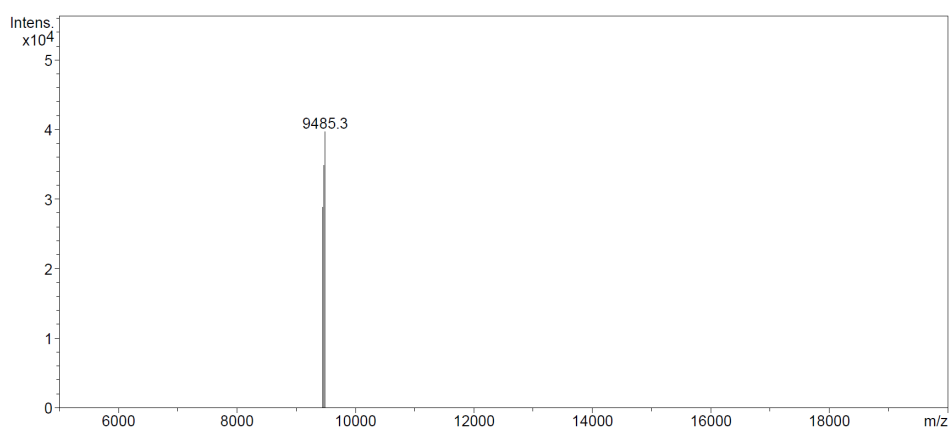


Figure S28: Confirmation of tagging reaction of ¹⁵N labelled ubiquitin S57C with Tb-M7-Nitro monitored by ESI-MS.

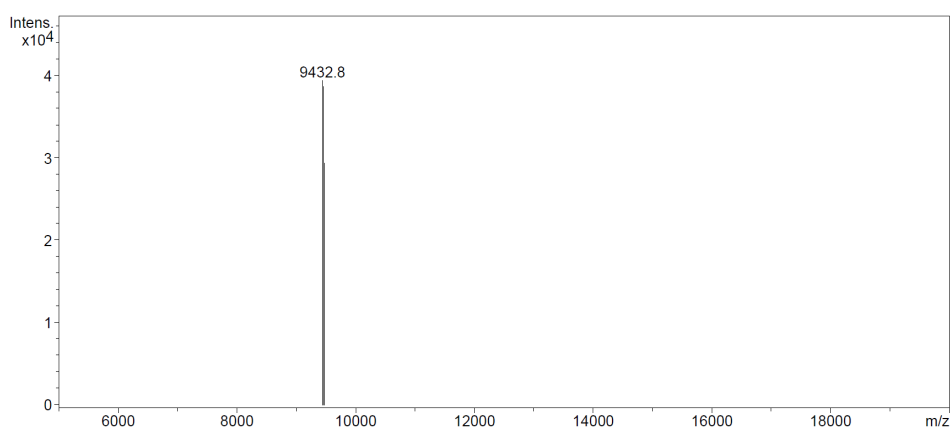


Figure S29: Confirmation of tagging reaction of ¹⁵N labelled ubiquitin S57C with Lu-M7-Nitro monitored by ESI-MS.

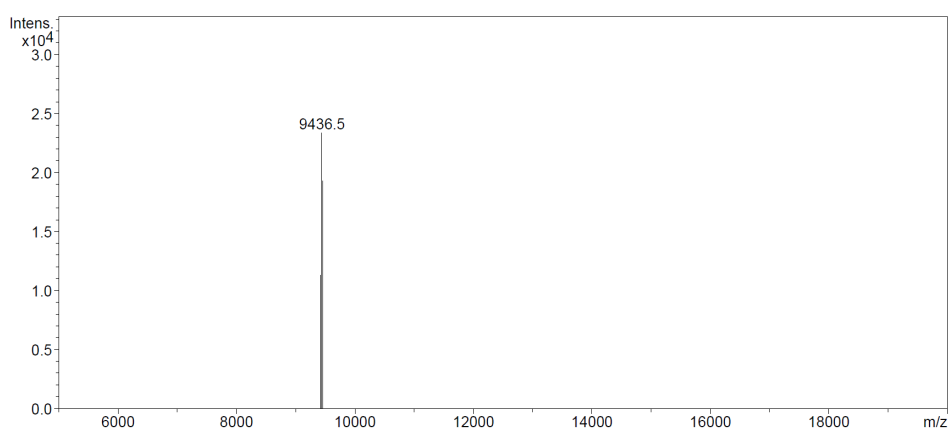


Figure S30: Confirmation of tagging reaction of ¹⁵N labelled ubiquitin S57C with Gd-M7-Nitro monitored by ESI-MS.

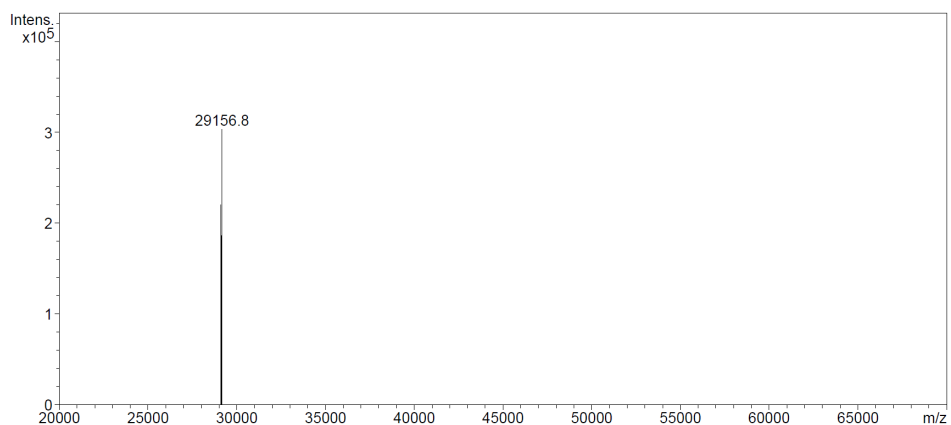


Figure S31: ESI-MS spectrum of selectively ^{15}N leucine labelled hCA II S50C.

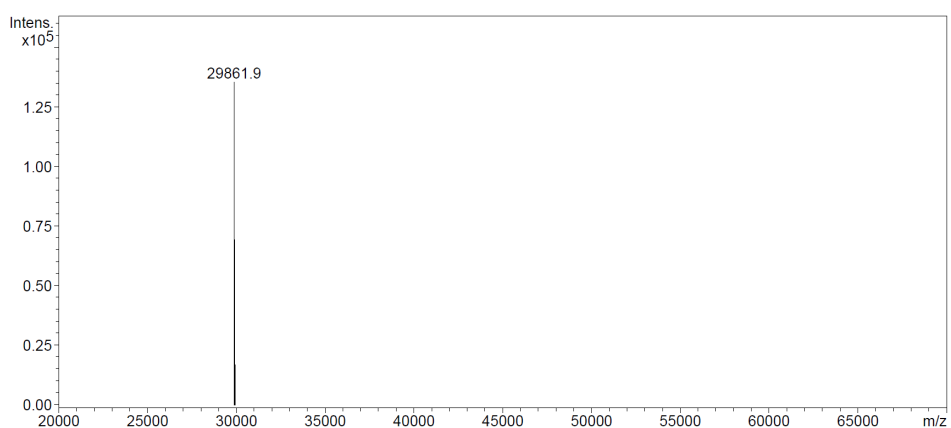


Figure S32: Confirmation of tagging reaction of selectively ^{15}N leucine labelled hCA II S50C with Tm-M7-Nitro monitored by ESI-MS.

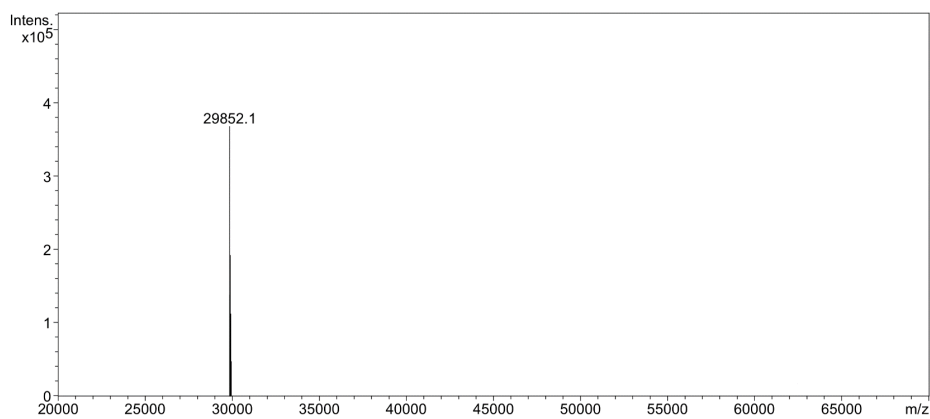


Figure S33: Confirmation of tagging reaction of selectively ^{15}N leucine labelled hCA II S50C with Dy-M7-Nitro monitored by ESI-MS.

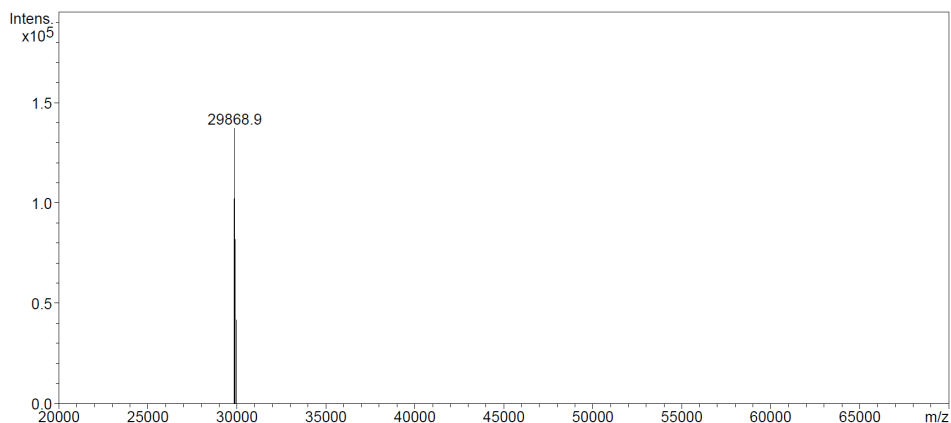


Figure S34: Confirmation of tagging reaction of selectively ¹⁵N leucine labelled hCA II S50C with Tb-M7-Nitro monitored by ESI-MS.

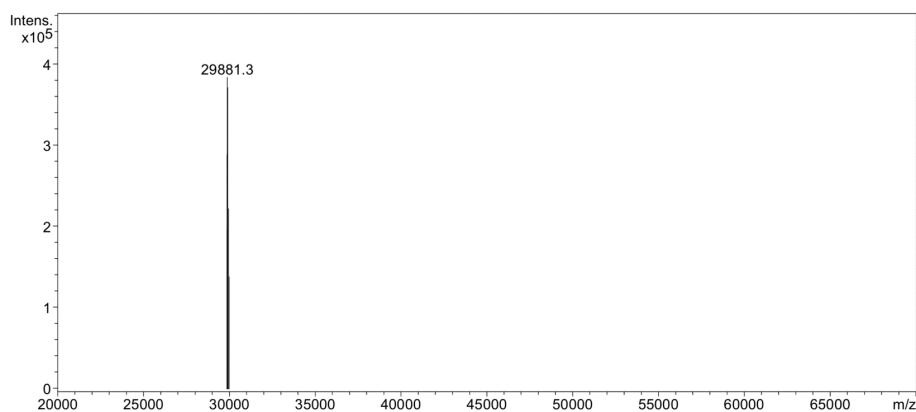


Figure S35: Confirmation of tagging reaction of selectively ¹⁵N leucine labelled hCA II S50C with Yb-M7-Nitro monitored by ESI-MS.

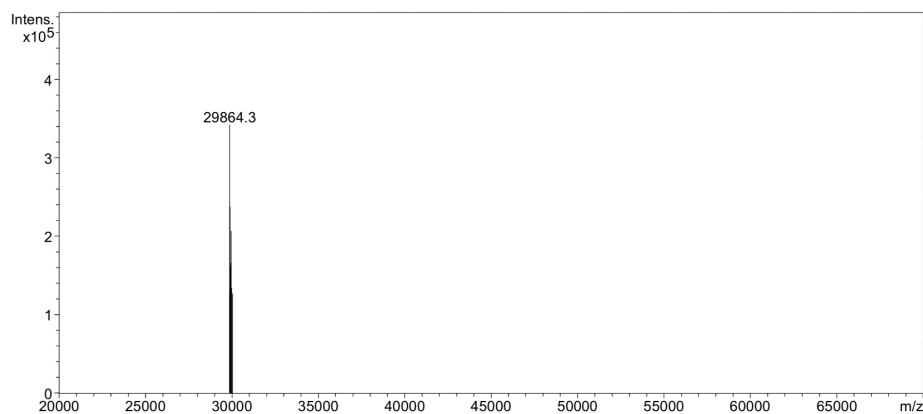


Figure S36: Confirmation of tagging reaction of selectively ¹⁵N leucine labelled hCA II S50C with Lu-M7-Nitro monitored by ESI-MS.

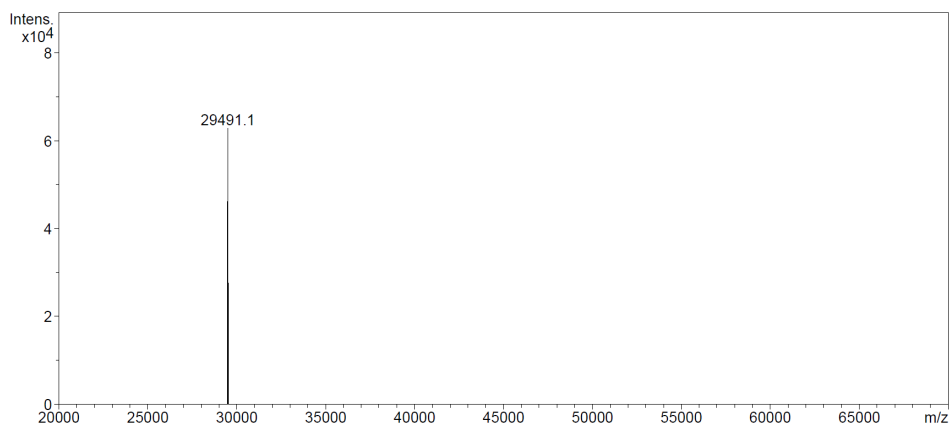


Figure S37: ESI-MS spectrum of uniformly ^{15}N labelled hCA II S50C.

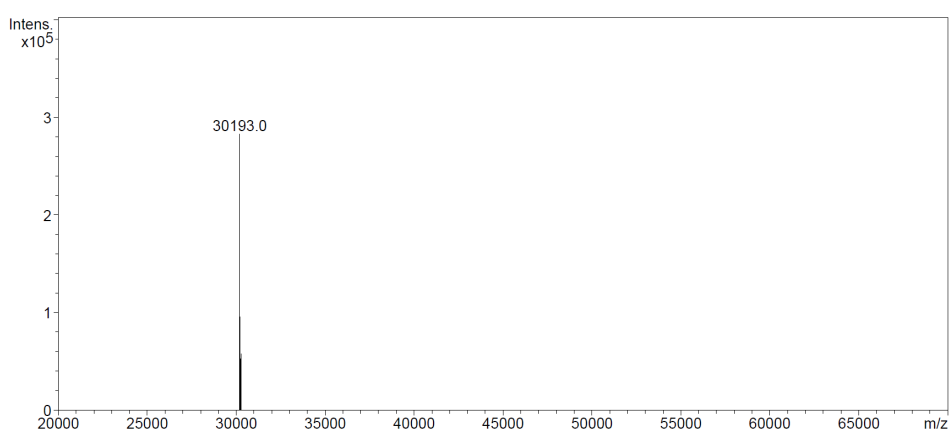


Figure S38: Confirmation of tagging reaction of uniformly ^{15}N labelled hCA II S50C with Lu-M7-Nitro monitored by ESI-MS.

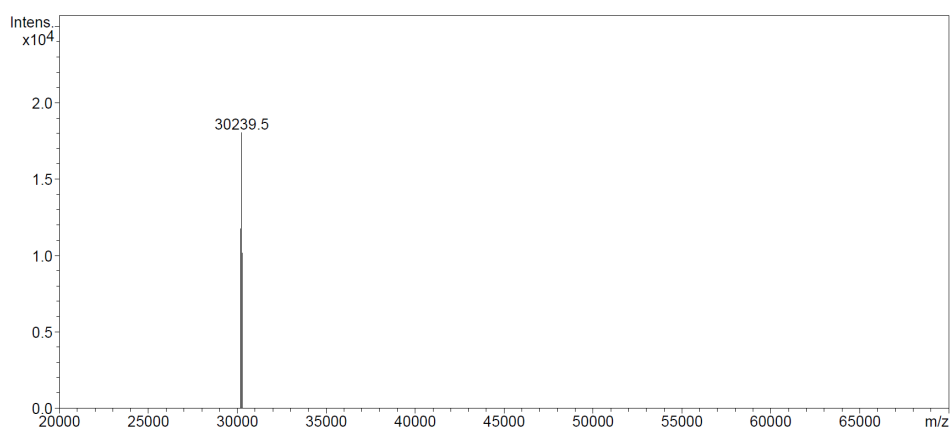


Figure S39: Confirmation of tagging reaction of uniformly ^{15}N labelled hCA II S50C with Gd-M7-Nitro monitored by ESI-MS.

Complete fitting results using Numbat, Paramagpy and Fanten

Table S17: Complete fitting results using Numbat⁸, Paramagpy¹⁰ and Fanten⁹. Induced axial and rhombic components of the paramagnetic susceptibility tensors ($\Delta\chi_{ax}$ and $\Delta\chi_{rh}$, in 10^{-32} m^3), metal position in PDB coordinate frame (X_{metal} , Y_{metal} , Z_{metal} , in Å), Euler angles (α , β , γ , in °) and quality factor (Q, mathematical definition in SI, p. 1) on ubiquitin S57C (pH 6.0) and selectively ¹⁵N leucine labelled hCA II S50C (pH 6.8) at 298 K. Note: Numbat⁸ and Paramagpy¹⁰ yield anisotropy parameters in UTR convention, Fanten⁹ depicts tensors in the Fanten convention. For RDC only fits, the metal position of the corresponding PCS fit was given.

Protein mutant + fitting routine	PDB	N° PCS	Ln ³⁺	$\Delta\chi_{ax}$ (10^{-32} m^3)	$\Delta\chi_{rh}$ (10^{-32} m^3)	X_{metal} (Å)	Y_{metal} (Å)	Z_{metal} (Å)	Distance to C _{8,cys} (Å)	α (°)	β (°)	γ (°)	Q (%)
UbiS57C Numbat	1UBI	32	Tm	60.9 ± 1.9	12.3 ± 1.3	16.3	16.9	12.3	6.9	110.9	70.4	101.4	5.1
		32	Dy	94.3 ± 2.2	10.7 ± 0.7					133.7	161.5	128.4	2.4
		44	Tb	65.9 ± 1.6	27.9 ± 0.8					128.0	162.4	110.4	7.0
UbiS57C Paramagpy	1UBI	32	Tm	73.3	17.9	15.8	16.1	11.9	7.4	112.4	68.2	100.6	5.1
		32	Dy	112.4	15.1					131.3	161.1	114.6	2.6
		44	Tb	79.5	34.9					125.2	162.2	104.6	7.2
UbiS57C Fanten	1UBI	32	Tm	64.5	-12.7	16.0	16.4	11.1	7.1	167.2	-112.6	112.1	6.1
		32	Dy	101.6	-9.7					153.2	-20.4	126.3	4.9
		44	Tb	72.8	-30.7					-171.8	-161.3	-61.6	10.1
UbiS57C Fanten jointly with RDC	1UBI	32	Tm	66.6	-12.5	14.2	17.4	13.4	9.2	164.0	-103.2	107.4	14.2
		32	Dy	106.7	-2.0					144.7	-164.8	-42.0	8.5
		44	Tb	78.8	-31.8					144	-15.4	144.5	19.3
hCA II S50C Numbat	3KS3	10	Tm	71.0 ± 1.8	6.2 ± 0.9	-31.0	4.4	14.9	7.1	174.7	17.4	139.0	4.3
		22	Dy	95.9 ± 1.1	13.1 ± 2.5					179.8	112.5	64.5	2.1
		38	Tb	63.6 ± 0.9	22.5 ± 1.4					4.9	62.5	109.5	2.0
		22	Yb	13.8 ± 0.6	7.5 ± 0.4					36.7	163.0	115.4	3.5
hCA II S50C Paramagpy	3KS3	10	Tm	89.2	13.5	-31.2	4.2	14.9	7.4	161.5	17.2	0.6	4.9
		22	Dy	98.2	17.3					1.1	67.5	114.9	2.5
		38	Tb	65.5	24.9					6.5	62.6	108.3	2.0
		22	Yb	14.4	7.7					38.4	162.0	115.8	3.6
hCA II S50C Fanten	3KS3	10	Tm	88.8	-21.6	-31.0	5.0	14.9	6.9	15.5	-121.0	-110.8	7.1
		22	Dy	97.4	-14.9					-17.7	-112.3	-2.5	4.5
		38	Tb	64.8	-22.4					-14.7	-116.9	2.6	3.1
		22	Yb	14.8	-7.1					14.3	-164.9	-156.7	4.6
hCA II S50C Fanten jointly with RDC	3KS3	10	Tm	102.5	-24.0	-31.6	6.5	15.4	6.6	10.2	-117.9	-111.0	7.4
		22	Dy	100.2	-23.1					20.6	-109.3	-8.0	4.5
		38	Tb	74.0	-23.4					-21.0	-112.7	-2.5	16.2
		22	Yb	14.1	-6.3					-125.6	-151.3	-123.6	14.0
Protein mutant + fitting routine	PDB	N° RDC	Ln ³⁺	$\Delta\chi_{ax}$ (10^{-32} m^3)	$\Delta\chi_{rh}$ (10^{-32} m^3)	X_{metal} (Å)	Y_{metal} (Å)	Z_{metal} (Å)	Distance to C _{8,cys} (Å)	α (°)	β (°)	γ (°)	Q (%)
UbiS57C Paramagpy	1UBI	17	Tm	59.7	11.8	15.8	16.1	11.9	7.4	107.4	76.8	105.7	18.0
		15	Dy	96.0	3.2					137.1	165.3	41.4	18.2
		21	Tb	71.3	29.3					146.2	164.3	126.8	20.5
UbiS57C Fanten	1UBI	17	Tm	66.3	-13.1	16.0	16.4	11.1	7.1	164.3	-103.2	107.4	18.0
		15	Dy	106.7	-3.5					-131.4	-14.7	137.1	18.2
		21	Tb	79.2	-32.5					143.2	-15.7	146.2	20.5
UbiS57C Fanten jointly with PCS	1UBI	17	Tm	66.6	-12.5	14.2	17.4	13.4	9.2	164.0	-103.2	107.4	18.1
		15	Dy	106.7	-19.9					144.7	-164.8	-42.0	18.3
		21	Tb	78.8	-31.8					144	-15.4	144.5	20.5
hCA II S50C Paramagpy	3KS3	18	Tb	67.4	21.0	-31.2	4.2	14.9	7.4	176.7	111.9	66.8	18.3
		11	Yb	12.5	5.4					56.7	151.1	143.6	50.1
hCA II S50C Fanten	3KS3	18	Tb	74.9	-23.3	-31.0	5.0	14.9	6.9	-23.2	-111.9	-3.3	18.3
		11	Yb	13.9	-6.0					-126.4	-151.1	-123.3	50.1
hCA II S50C Fanten jointly with PCS	3KS3	18	Tb	74.0	-23.4	-31.6	6.5	15.4	6.6	-21.0	-112.7	-2.5	18.5
		11	Yb	14.1	-6.3					-125.6	-151.3	-123.6	50.1

Fitting results using Fanten (PCS only, RDC only and PCS/RDC combined fits)

Table S18: Fitting results using Fanten⁹ (PCS only and PCS/RDC combined fit). Induced axial and rhombic components of the paramagnetic susceptibility tensors ($\Delta\chi_{ax}$ and $\Delta\chi_{rh}$, in 10^{-32} m^3), metal position in PDB coordinate frame (X_{metal} , Y_{metal} , Z_{metal} , in Å), Euler angles (α , β , γ , in °) and quality factor (Q, mathematical definition in SI, p. 1) on ubiquitin S57C (pH 6.0) and selectively ¹⁵N leucine labelled hCA II S50C (pH 6.8) at 298 K.

Protein mutant + fitting routine	PDB	N° PCS	Ln ³⁺	$\Delta\chi_{ax}$ (10^{-32} m^3)	$\Delta\chi_{rh}$ (10^{-32} m^3)	X_{metal} (Å)	Y_{metal} (Å)	Z_{metal} (Å)	Distance to C _β ,Cys (Å)	α (°)	β (°)	γ (°)	Q (%)
UbiS57C Fanten; PCS only	1UBI	32	Tm	64.5	-12.7	16.0	16.4	11.1	7.1	167.2	-112.6	112.1	6.1
		32	Dy	101.6	-9.7					153.2	-20.4	126.3	4.9
		44	Tb	72.8	-30.7					-171.8	-161.3	-61.6	10.1
UbiS57C Fanten; PCS jointly with RDC	1UBI	32	Tm	66.6	-12.5	14.2	17.4	13.4	9.2	164.0	-103.2	107.4	14.2
		32	Dy	106.7	-2.0					144.7	-164.8	-42.0	8.5
		44	Tb	78.8	-31.8					144	-15.4	144.5	19.3
hCA II S50C Fanten; PCS only	3KS3	10	Tm	88.8	-21.6	-31.0	5.0	14.9	6.9	15.5	-121.0	-110.8	7.1
		22	Dy	97.4	-14.9					-17.7	-112.3	-2.5	4.5
		38	Tb	64.8	-22.4					-14.7	-116.9	2.6	3.1
		22	Yb	14.8	-7.1					14.3	-164.9	-156.7	4.6
hCA II S50C Fanten; PCS jointly with RDC	3KS3	10	Tm	102.5	-24.0	-31.6	6.5	15.4	6.6	10.2	-117.9	-111.0	7.4
		22	Dy	100.2	-23.1					20.6	-109.3	-8.0	4.5
		38	Tb	74.0	-23.4					-21.0	-112.7	-2.5	16.2
		22	Yb	14.1	-6.3					-125.6	-151.3	-123.6	14.0
UbiS57C Fanten; RDC only	1UBI	17	Tm	66.3	-13.1	16.0	16.4	11.1	7.1	164.3	-103.2	107.4	18.0
		15	Dy	106.7	-3.5					-131.4	-14.7	137.1	18.2
		21	Tb	79.2	-32.5					143.2	-15.7	146.2	20.5
UbiS57C Fanten; RDC jointly with PCS	1UBI	17	Tm	66.6	-12.5	14.2	17.4	13.4	9.2	164.0	-103.2	107.4	18.1
		15	Dy	106.7	-19.9					144.7	-164.8	-42.0	18.3
		21	Tb	78.8	-31.8					144	-15.4	144.5	20.5
hCA II S50C Fanten; RDC	3KS3	18	Tb	74.9	-23.3	-31.0	5.0	14.9	6.9	-23.2	-111.9	-3.3	18.3
		11	Yb	13.9	-6.0					-126.4	-151.1	-123.3	50.1
hCA II S50C Fanten; RDC jointly with PCS	3KS3	18	Tb	74.0	-23.4	-31.6	6.5	15.4	6.6	-21.0	-112.7	-2.5	18.5
		11	Yb	14.1	-6.3					-125.6	-151.3	-123.6	50.1

Note: Fanten⁹ depicts tensors not in UTR but in the Fanten convention.

PCS data were fitted for each construct, using all metals simultaneously, to yield a common metal position.

For RDC only fits, the metal position of the corresponding PCS fit was given.

The joint PCS/RDC fits give only one common set of tensor parameters and metal coordinates, but separate Q factors for PCS and RDC.

Comparison of ^1H line width of Leu-141 in hCA II S50C tagged with Lu-M7Nitro (red) or Tm-M7Nitro (blue)

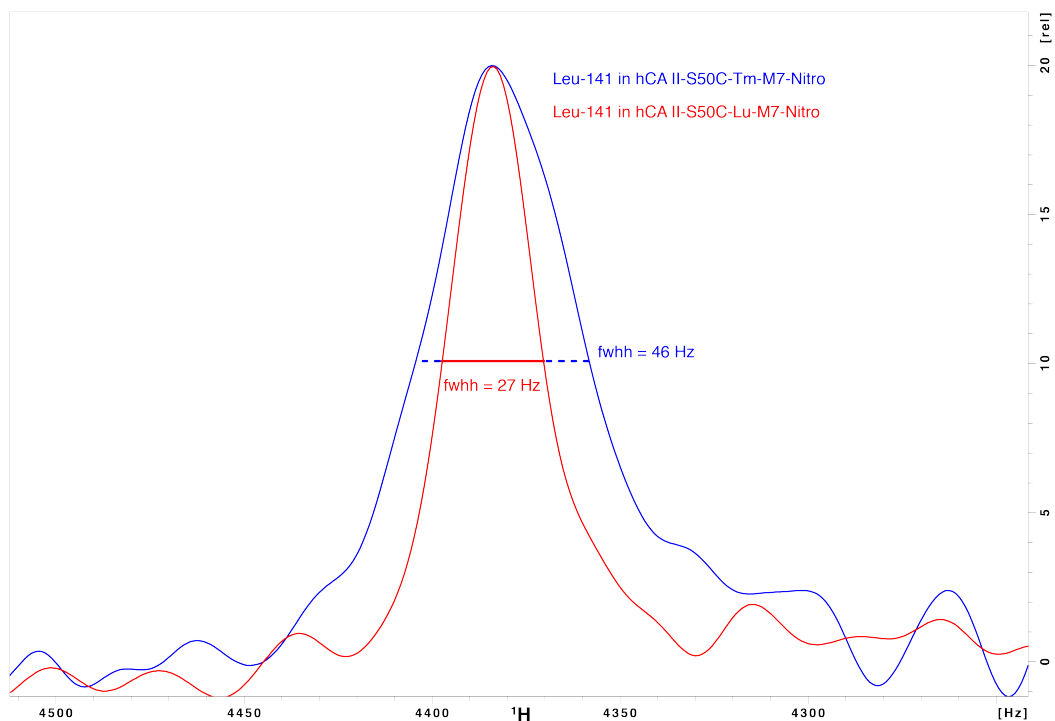


Figure S40: Comparison of the ^1H line width of a representative ^1H - ^{15}N -HSQC cross peak (Leu-141) in selectively ^{15}N leucine labelled hCA II S50C tagged with Lu-M7Nitro (red) or Tm-M7Nitro (blue). The difference in line width caused predominantly by Curie line broadening amounts to ca. 20 Hz @ 600 MHz. The amide proton of Leu-141 is located 23.6 Å away from the metal position. The chemical shift offset and the intensity of the Lu-tagged spectrum were adjusted to enable direct comparison.

Comparison of tagging kinetics of Ln-M7Nitro, Ln-M7PyThiazole and M7-FPy

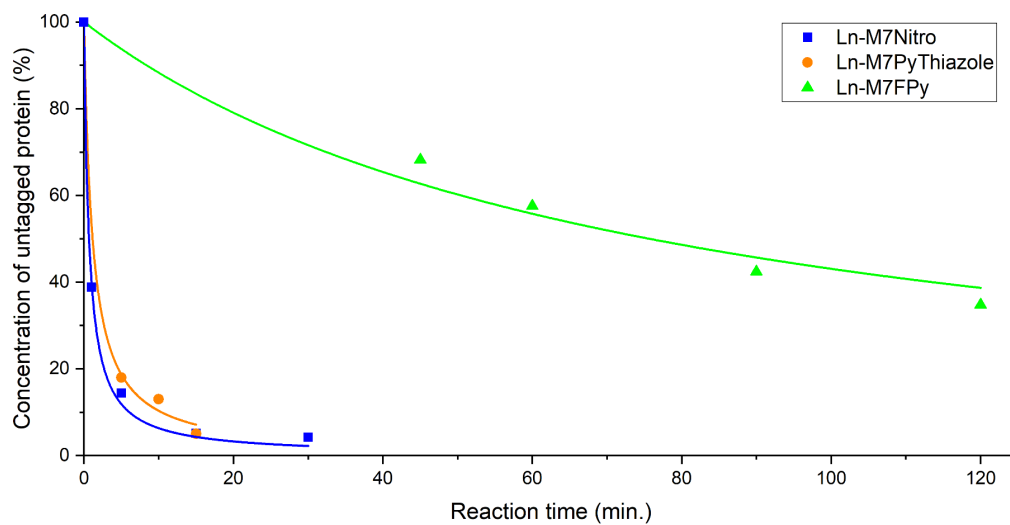


Figure S41: Comparison of tagging kinetics of Ln-M7Nitro, Ln-M7PyThiazole and Ln-M7-FPy with ubiquitin S57C. Ln-M7Nitro shows the fastest conversion of monomeric ubiquitin and tagging is complete after 30 min at rt in 10 mM phosphate buffer with pH 7.0.

References

1. Battiste, J. L.; Wagner, G. *Biochemistry* **2000**, *39*, 5355-5365.
2. Müntener, T.; Kottelat, J.; Huber, A.; Häussinger, D. *Bioconjugate Chem.* **2018**, *29*, 3344-3351.
3. Sass, J.; Cordier, F.; Hoffmann, A.; Rogowski, M.; Cousin, A.; Omichinski, J. G.; Löwen, H.; Grzesiek, S. *J. Am. Chem. Soc.* **1999**, *121*, 2047-2055.
4. Varghese, S.; Halling, P. J.; Häussinger, D.; Wimperis, S. *J. Phys. Chem. C* **2016**, *120*, 28717-28726.
5. Vranken, W. F.; Boucher, W.; Stevens, T. J.; Fogh, R. H.; Pajon, A.; Llinas, M.; Ulrich, E. L.; Markley, J. L.; Ionides, J.; Laue, E. D. *Proteins: Struct., Funct., Bioinf.* **2005**, *59*, 687-696.
6. Ramage, R.; Green, J.; Muir, T. W.; Ogunjobi, O. M.; Love, S.; Shaw, K. *Biochem. J* **1994**, *299* (Pt 1), 151-8.
7. Avvaru, B. S.; Kim, C. U.; Sippel, K. H.; Gruner, S. M.; Agbandje-McKenna, M.; Silverman, D. N.; McKenna, R. *Biochemistry* **2010**, *49*, 249-51.
8. Schmitz, C.; Stanton-Cook, M. J.; Su, X.-C.; Otting, G.; Huber, T. *J. Biomol. NMR* **2008**, *41*, 179.
9. Rinaldelli, M.; Carlon, A.; Ravera, E.; Parigi, G.; Luchinat, C. *J. Biomol. NMR* **2015**, *61*, 21-34.
10. Orton, H. W.; Huber, T.; Otting, G. *Magn. Reson.* **2020**, *1*, 1-12.
11. Keizers, P. H.; Saragliadis, A.; Hiruma, Y.; Overhand, M.; Ubbink, M. *J. Am. Chem. Soc.* **2008**, *130*, 14802-14812.

4 Part II

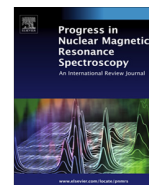
4.1 Design and application of lanthanide chelating tags for pseudocontact shift nuclear magnetic resonance spectroscopy on biomacromolecules

Reproduced from D. Joss, D. Häussinger, *Progress in Nuclear Magnetic Resonance Spectroscopy*, **2019**, 114-115, 284-312 with permission of Elsevier (<https://doi.org/10.1016/j.pnmrs.2019.08.002>).



Contents lists available at ScienceDirect

Progress in Nuclear Magnetic Resonance Spectroscopy

journal homepage: www.elsevier.com/locate/pnmrs

Design and applications of lanthanide chelating tags for pseudocontact shift NMR spectroscopy with biomacromolecules



Daniel Joss, Daniel Häussinger*

University of Basel, St. Johannis-Ring 19, 4056 Basel, Switzerland

Edited by Geoffrey Bodenhausen and David Neuhaus

ARTICLE INFO

Article history:

Received 31 July 2019

Accepted 24 August 2019

Available online 28 August 2019

Keywords:

Paramagnetic NMR

Pseudocontact shift

Residual dipolar coupling

Lanthanide chelating tag

Biomacromolecule

ABSTRACT

In this review, lanthanide chelating tags and their applications to pseudocontact shift NMR spectroscopy as well as analysis of residual dipolar couplings are covered. A complete overview is presented of DOTA-derived and non-DOTA-derived lanthanide chelating tags, critical points in the design of lanthanide chelating tags as appropriate linker moieties, their stability under reductive conditions, e.g., for in-cell applications, the magnitude of the anisotropy transferred from the lanthanide chelating tag to the biomacromolecule under investigation and structural properties, as well as conformational bias of the lanthanide chelating tags are discussed. Furthermore, all DOTA-derived lanthanide chelating tags used for PCS NMR spectroscopy published to date are displayed in tabular form, including their anisotropy parameters, with all employed lanthanide ions, C_B-Ln distances and tagging reaction conditions, i.e., the stoichiometry of lanthanide chelating tags, pH, buffer composition, temperature and reaction time. Additionally, applications of lanthanide chelating tags for pseudocontact shifts and residual dipolar couplings that have been reported for proteins, protein-protein and protein-ligand complexes, carbohydrates, carbohydrate-protein complexes, nucleic acids and nucleic acid-protein complexes are presented and critically reviewed. The vast and impressive range of applications of lanthanide chelating tags to structural investigations of biomacromolecules in solution clearly illustrates the significance of this particular field of research. The extension of the repertoire of lanthanide chelating tags from proteins to nucleic acids holds great promise for the determination of valuable structural parameters and further developments in characterizing intermolecular interactions.

© 2019 Elsevier B.V. All rights reserved.

Contents

1. Introduction	285
1.1. Paramagnetic nuclear magnetic resonance spectroscopy of biomacromolecules	285
1.2. Pseudocontact shifts	286
1.3. Residual dipolar couplings	287
2. Lanthanide chelating tags	287
2.1. Introduction & overview	287
2.2. Non-DOTA-based lanthanide chelating tags	288
2.3. DOTA-based lanthanide chelating tags	291
2.4. Tabular overview of DOTA-derived LCTs	295
2.5. Parameters in the design of LCTs	295
3. Applications of lanthanide chelating tags to the structural characterization of biomacromolecules	301
3.1. Applications of lanthanide chelating tags to proteins I: pseudocontact shifts	301
3.2. Applications of lanthanide chelating tags to proteins II: residual dipolar couplings	303
3.3. Applications of lanthanide chelating tags to proteins III: protein-ligand complexes	304

* Corresponding author.

E-mail addresses: daniel.joss@unibas.ch (D. Joss), daniel.haeussinger@unibas.ch (D. Häussinger).<https://doi.org/10.1016/j.pnmrs.2019.08.002>

0079-6565/© 2019 Elsevier B.V. All rights reserved.

3.4. Applications of lanthanide chelating tags to carbohydrates.....	305
3.5. Applications of lanthanide chelating tags to nucleic acids.....	307
4. Conclusion & outlook.....	308
Declaration of competing Interest.....	308
Acknowledgements.....	308
References.....	308

1. Introduction

1.1. Paramagnetic nuclear magnetic resonance spectroscopy of biomacromolecules

Nuclear magnetic resonance (NMR) spectroscopy is one of the most important analytical techniques in many different fields of chemistry and biology, from organic synthesis to structural biology. In chemistry, NMR is mainly used as a routine tool for confirming the topology of small molecules as well as 3D structure elucidation. The development of more advanced NMR techniques, i.e., homo- and heteronuclear NMR pulse sequences for establishing through-space and through-bond correlations between nuclei that are suitable for complex structural studies, turned out to be crucial for the understanding of the structure and function of biomacromolecules, e.g. of proteins as well as deoxyribonucleic acids (DNAs) and ribonucleic acids (RNAs). The limitation of nuclear Overhauser effect spectroscopy (NOESY) to distances below approx. 5–8 Å and the lack of specificity of chemical shift perturbation (CSP) experiments, however, explain that long-range paramagnetic effects are highly valuable, complementary tools to measure distances and orientational restraints in biomacromolecules [1–24]. Paramagnetic effects observed in NMR include paramagnetic shifts, i.e., contact (CS) and pseudocontact shifts (PCS) [1–3,6,8,10,14,23–31], paramagnetic relaxation enhancement (PRE) [16,17,22,32–36], residual dipolar couplings (RDC) [37–45], dipolar shift anisotropy (DSA), cross-correlations between DSA and dipolar relaxation (DSA/DD) and residual anisotropic chemical shifts (RACS). The different effects associated with the introduction of a paramagnetic lanthanide ion yield comple-

mentary structural restraints and therefore offer a powerful toolbox for the structure determination of biomacromolecules in solution (Fig. 1). Whereas through-bond CS effects can be neglected for studies of biomacromolecules, PCSs yield structural restraints for distances up to more than 100 Å due to their favourable r^{-3} distance dependence and the 3D spatial information that is encoded in the induced shifts. PREs offer an opportunity to obtain information on the distance between a nucleus and a paramagnetic centre by line broadening and hence decrease in signal intensity. The r^{-6} distance dependence of PRE effects is less favourable than for PCSs (r^{-3}) and no angular information can be obtained, but the magnitude of the observed effects suffers less from motional averaging. RDCs provide valuable information about the relative orientation of bond vectors in space. To observe RDCs, a partial alignment of the protein is required. This can be obtained either by the introduction of a paramagnetic centre or by the use of an alignment medium, such as bacteriophages, bicelles or acrylamide gels [46,47]. Because of the favourable r^{-3} distance dependence of PCSs, and the lack of distance dependence for RDCs (although motional averaging can lead to decreased RDCs at larger distances from the paramagnetic centre), as well as the fact that PCSs provide not only distance but also angular information and thus yield valuable positional restraints, this review focuses on paramagnetic tags that allow one to exploit PCSs and RDCs for the structural investigation of biomacromolecules and their complexes.

Soon after a landmark study published in 1971 by Barry et al. [49], who could quantitatively determine the conformations of mononucleotides in solution by scanning through molecular geometries and matching them with the lanthanide induced NMR shifts using a Ferranti Argus 500 computer, the theoretical

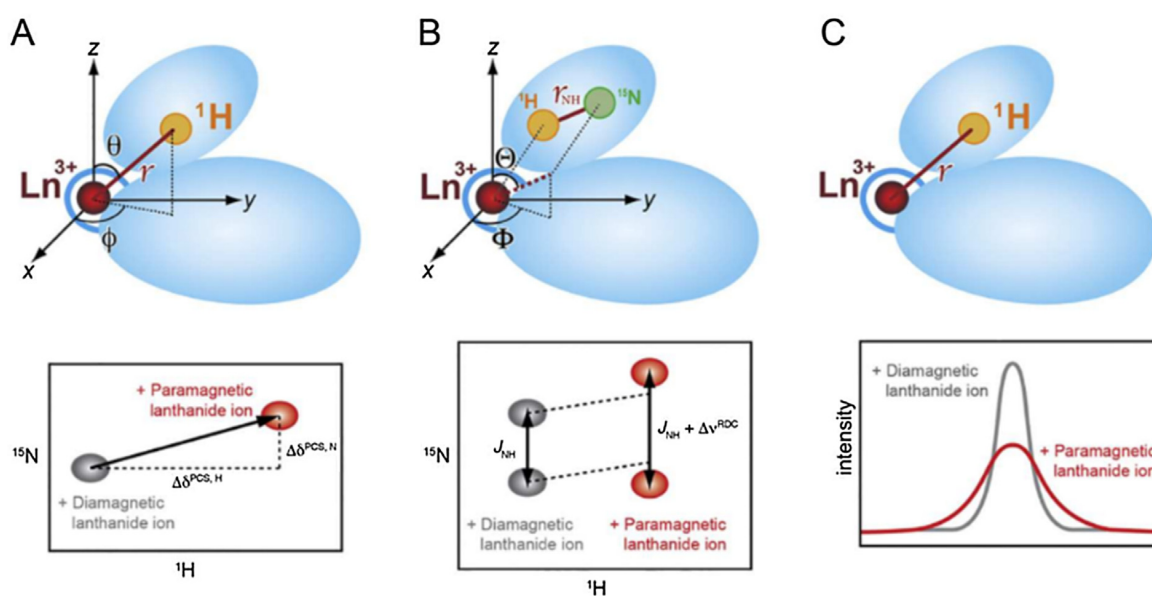


Fig. 1. Effects associated with the introduction of a lanthanide ion [48]. (A) 3D spatial information obtained by PCS. (B) Orientational restraints obtained by RDCs. (C) Distance measurement using PRE.

background of the NMR shifts induced by paramagnetic lanthanides was established in the 1970s by Bleaney et al. [50,51]. However, criticisms and new findings arose recently concerning the validity of Bleaney's theory about the magnetic properties of lanthanide complexes in the context of paramagnetic NMR [52,53]. The very first applications of paramagnetic NMR include europium shift reagents that allow one to increase the shift range in order to overcome signal overlap and determine enantiomeric ratios of chiral substances by forming diastereomeric complexes in-situ [54–58]. After the successful exploitation of paramagnetic effects of metal ions contained in proteins by Bertini et al., e.g., in zinc finger moieties [16], and the determination of the associated paramagnetic susceptibility tensors, the next step in the development of paramagnetic NMR led to lanthanide binding peptides (LBPs) with micro- and nanomolar affinity to lanthanides mimicking the zinc finger motif in order to incorporate paramagnetic [15,59] and fluorescent [60,61] metal ions into proteins without native metal binding sites. Subsequently, the development of single- and double-armed lanthanide chelating tags (LCTs) that can be attached to a protein of interest and, thus, yield valuable structural restraints for the determination of the structures of biomacromolecules in solution, strongly expanded the field of paramagnetic NMR [9,13,14,24,62,63]. Today, paramagnetic NMR is an established method for structural biology and includes many different applications, such as refining crystal structures of proteins [64], probing solution dynamics of proteins [65], localization of ligands attached to proteins [20,66–71] (Fig. 2), study of protein-protein complexes [72], alleviation of signal dispersion in crowded ^1H - ^{15}N HSQC spectra [73], facilitating analysis of large proteins [12], conformational analysis of carbohydrates [74], as well as investigation of the solution structure of nucleic acids [30].

Besides applications of paramagnetic effects to protein NMR spectroscopy, detailed studies were performed of the solvent dependence, as well as pH and temperature effects on the magnetic anisotropy caused by paramagnetic molecules, either alone or attached to proteins [28,75].

Furthermore, computational approaches, such as density functional theory (DFT) calculations, allow one to predict the structure and conformational bias of lanthanide complexes *in vacuo* and in solution as well as anisotropy tensors for electron paramagnetic resonance (EPR) studies and are therefore under current investigation in order to facilitate the use of paramagnetic effects by decreasing the experimental work required for the development of new LCTs [23,76–83]. Very recent work proves the semi-empirical approach to be superior in reproducing PCS compared to quantum chemistry approaches [84].

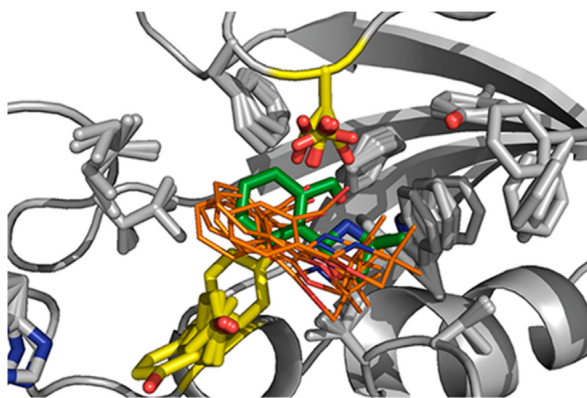


Fig. 2. Localization of a ligand attached to a protein: green: averaged NOESY structure, orange: five best structures obtained by PCS-fitting [69]. (Reproduced with permission from Ref. [69]. Copyright 2013 American Chemical Society.)

Comprehensive reviews about lanthanide chelating tags and their applications are available in the literature [1,5,7,48,85–89]. A very recent article by Su et al. documents his contributions to the field [90]. Furthermore, reviews were published that give an insight into strategies for measurements of PCSs in proteins [91], prospects for lanthanides in biomolecular NMR using LCTs [92], analysis of protein-ligand [20] as well as protein-protein complexes [93,94] using PCSs and RDCs, the combination of long-range paramagnetic restraints and crystallography [95,96], sparse-labelling in combination with long-range restraints obtained by paramagnetic NMR [97] and LCTs used for the conformational analysis and interactions of oligosaccharides [98]. This review will focus on recent developments in the synthesis of LCTs and their application for PCS and RDC to structural biology, with an emphasis on DOTA-based LCTs. As the intricate fundamental physics of PCS and RDC have been discussed several times in great detail [85,99,100], we will merely summarize the basics in the following sections.

1.2. Pseudocontact shifts

The dipolar, through-space effect on the shifts of signals in NMR due to the introduction of a paramagnetic centre is termed pseudocontact shift [99]. The shift perturbation arises from a contribution of the unpaired electrons of the paramagnetic centre to the effective magnetic field that a specific nucleus experiences. This shift perturbation depends on the distance and angular position of the nucleus with respect to the paramagnetic centre. Considering this 3D spatial information and the r^{-3} distance dependence of the effect, valuable structural information can be gained about the molecule in the vicinity of the paramagnetic centre. The shift that a nucleus experiences can be calculated from the following equations that can be expressed in polar or in Cartesian coordinates [99]:

PCS (polar coordinates)

$$\delta^{\text{PCS}} = \frac{1}{12\pi r^3} \left[\Delta\chi_{ax}(3\cos^2\theta - 1) + \frac{3}{2}\Delta\chi_{rh}\sin^2\theta\cos 2\phi \right] \quad (1)$$

PCS (cartesian coordinates)

$$\delta^{\text{PCS}}(x, y, z) = \frac{1}{12\pi r^3} \left[\Delta\chi_{ax} \frac{2z^2 - x^2 - y^2}{r^2} + \frac{3}{2}\Delta\chi_{rh} \frac{x^2 - y^2}{r^2} \right] \quad (2)$$

Eight parameters are required to define the tensor that describes the anisotropic part of the magnetic susceptibility associated with a paramagnetic centre, e.g., of an LCT when attached to a protein. Besides the x , y and z coordinates of the metal centre and the three Euler angles α , β and γ that describe the orientation of the tensor with regard to the protein, the ellipticity of the magnetic susceptibility is described by the axial component and the shape of the tensor by the ratio of the axiality to the rhombicity of the tensor. The axiality and rhombicity of a tensor are defined as follows [99]:

$$\text{Axiality} \quad \Delta\chi_{ax} = \chi_{zz} - \frac{\chi_{xx} + \chi_{yy}}{2} \quad (3)$$

$$\text{Rhombicity} \quad \Delta\chi_{rh} = \chi_{xx} - \chi_{yy} \quad (4)$$

As depicted in Fig. 3, anisotropic χ -tensors are usually displayed as isosurfaces for given PCS values. For a pronounced axial/rhombic ratio, a tensor resembling the shape of a d_{zz} orbital is obtained, whereas for a more balanced axial/rhombic ratio a tensor shape resembling the shape of a d_{xy} orbital is obtained.

In order to study proteins by PCS in NMR spectroscopy, ^1H - ^{15}N HSQC experiments with both diamagnetic and paramagnetic samples have to be performed. Based on the PCSs observed for the

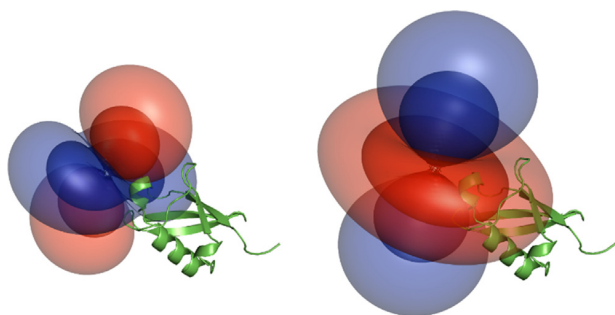


Fig. 3. Isosurfaces of the anisotropy parameters generated by the dysprosium (left) and thulium (right) complex of Ln-DOTA-M8-(4R4S)-SSPy when attached to ubiquitin S57C (inner layer: 1.0 ppm, outer layer: 0.25 ppm). While Dy^{3+} induces a tensor with a reasonably balanced axial/rhombic ratio ($\Delta\chi_{ax}/\Delta\chi_{rh} = 2.1$), Tm^{3+} causes a tensor with a pronounced axial/rhombic ratio ($\Delta\chi_{ax}/\Delta\chi_{rh} = 6.5$) [23].

paramagnetic sample compared to chemical shifts of the diamagnetic sample, $\Delta\chi$ -parameters can be obtained by application of a Monte-Carlo simulation and the crystal structure can be refined in solution [101]. In most cases, the pseudocontact shift contribution, when expressed in units of ppm, is equal in both dimensions of the spectrum, so that the shifted signals are found at an angle of 45° relative to the corresponding diamagnetic signal. Due to the common distance between the amide ^{15}N and the $^1H^N$ hydrogen of $\sim 1.0 \text{ \AA}$ for most residues [102], exceptions of this behaviour arise for N-H groups located on nodal planes of the anisotropy tensor or in regions with a very steep gradient. The resulting spectra are iteratively assigned and the obtained shifts fitted to an already solved X-ray structure or a model based on the interpretation of chemical shifts using the software package Rosetta.

PCSs give rise to multiple applications from structure refinement to localization of ligands [19,20,66–71]. Deviations from the shifts that can be back-calculated from the crystal structure can be used to identify flexible regions in the protein. Furthermore, a protein with a paramagnetic centre incorporated allows one to perform, in combination with Rosetta, more elaborate predictions of the protein structure [103]. The required input for the structure determination of a protein is then reduced to PCS data for all amino $^1H^N$ protons and the amino-acid sequence of the protein. Based on a database containing structures for short oligopeptide fragments combined with restraints given by the dataset of PCSs, a *de novo* structure determination of a biomacromolecule can be achieved. Besides their use as structural restraints for the investigation of the structure of proteins, PCSs can also be exploited for the localization of ligands on a protein of interest [20,66–71]. By determination of the $\Delta\chi$ -parameters from multiple tagging sites, the ligand bound to the protein can then be localized based on its PCS. Among other examples, this localization method was demonstrated for 1H PCS of weakly bound ligands [69], for ^{19}F PCS detected via an one-dimensional ^{19}F PCS-CEST approach for weakly binding ligands in the intermediate exchange regime [66], as well as for ^{19}F -PCS of strongly binding ligands [71].

1.3. Residual dipolar couplings

RDCs can be observed in NMR experiments when a partial alignment leads to an incomplete averaging of anisotropic magnetic interactions and a dipole-dipole coupling between nuclear spins under investigation, e.g., between 1H and ^{15}N in N-H bonds in proteins, leads to a splitting of signals [46]. If the protein tumbles isotropically in solution, this results in an averaging to zero of the dipolar interactions. Partial alignment is therefore required for dipolar couplings to be observed. This can be achieved by either

introducing a paramagnetic centre covalently attached to the protein, thus leading to a field-dependent alignment, or by the use of alignment media such as polyacrylamide or collagen gels, leading to a field-independent alignment. Using RDCs, as in the case of PCSs, based on a known or predicted structure, e.g., from a Rosetta calculation, the $\Delta\chi$ -parameters can be fitted and a refined structure can be obtained. Since the orientation of most N-H bond vectors of the protein can be obtained by RDCs, the information is complementary to NOESY experiments and leads in combination with NOESY restraints to a more precise picture of the solution structure of the investigated biomacromolecule [104,105]. In contrast to PCSs (r^{-3}) or PREs (r^{-6}), RDCs are independent of the lanthanide – spin pair distance.

Dipolar coupling between two spins i and j : D_{ij}

$$= k_{ij} \left[A_d (3 \cos^2 \theta - 1) + \frac{3}{2} A_r \sin^2 \theta \cos(2\phi) \right] \quad (5)$$

$$\text{With } k_{ij} = -\frac{\gamma_i \gamma_j \mu_0 h}{16\pi^3 r_{ij}^3} \quad (6)$$

DOTA-based LCTs, e.g., CLaNP-5 [24] or DOTA-M8-(8S)-SSPy [9], can yield such valuable structural restraints. As recently demonstrated by Müntener et al., using a DOTA-based tag bearing a pyridine linker stable to reduction, RDCs up to 27.7 Hz were observed in *in-vitro* and in *in-cell* experiments [3] (Fig. 4). These valuable orientation restraints were then used to obtain a *de novo* structure for GB1 in combination with the Rosetta database. The obtained structure closely resembles the X-ray structure of the protein. For proteins for which an X-ray structure cannot be obtained, or when deviations between the structure in solution and the structure in crystalline state are expected, RDCs represent a valuable source of structural restraints.

RDCs are most valuable as a source of long-range structural restraints and can be used to obtain insights into specific biological problems, as, e.g., the determination of the relative orientation or exact angle between two helices in proteins, the extent of bending of a long helix in a protein, and the relative orientation of protein domains or subunits of a protein or protein-protein complex [106]. The combination of RDCs with other methods such as X-ray crystallography can lead to highly refined solution structures of both proteins and nucleic acids, e.g., of a protein-protein complex for which the structures of the individual subunits have been determined previously by X-ray crystallography. Less common applications of RDCs include the determination of the stereochemistry of natural products [107]. Besides the use of RDCs for structure determination of proteins, the helix curvature, subunit or domain orientation and stoichiometry of homomultimeric nucleic acid complexes were also investigated [108].

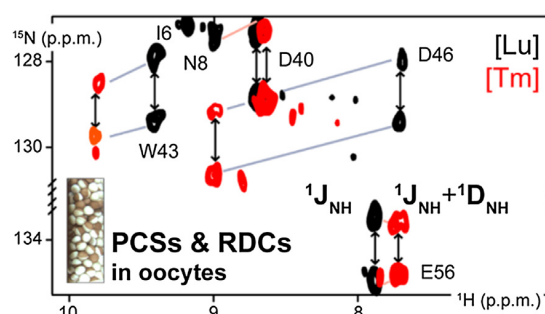


Fig. 4. RDCs measured on GB1 conjugated to an LCT in *Xenopus laevis* oocytes acquired at 600 MHz proton frequency [3]. (Reproduced with permission from Ref. [3]. Copyright 2016 American Chemical Society.)

To conclude, RDCs provide valuable long-range structural restraints and are able to drive the field of structural biology to a new level of precision in the structure elucidation of biomacromolecules in solution.

2. Lanthanide chelating tags

2.1. Introduction & overview

Paramagnetic NMR using lanthanides started with the exploitation of intrinsic metal binding sites, particularly Ca^{2+} binding sites, since Ca^{2+} ions display a similar oxophilicity and ionic radius (112 pm [109]) compared to Ln^{3+} ions (Ce–Lu: 114–98 pm [109]). In a comprehensive study, Bertini et al. compared the full series of lanthanide ions in a similar coordination environment in the protein calbindin D_{9k} [110] (the tensors of the full lanthanide series are depicted in Ref. [20]). Upon successful substitution of calcium by a lanthanide ion, paramagnetic effects as PCS can be measured directly, as shown by Bertini et al. for Calbindin D_{9k} that possesses two calcium binding sites [111]. However, this approach obviously only works for proteins containing a metal binding site. In order to study other targets by paramagnetic NMR, artificial binding sites have to be introduced or attached to the protein of interest.

As early as 2000, Ma et al. reported an EF-hand motif engineered to be attached to the N-terminus of a membrane-associated protein and thereby showed that it is possible to introduce lanthanides into biomacromolecules without a native metal binding site via LBPs to exploit induced RDCs [112]. LBPs mimic EF hand and zinc finger motifs in order to incorporate lanthanides into proteins that do not have any natural metal binding site [60,61,112–115] (Fig. 5). LBPs are advantageous with regard to tagging efficiency since peptide fusion always gives rise to very high ligation yields. However, a significant drawback is that the incorporation is only possible on the N- or the C-terminus of the investigated protein to avoid severely altering the structure of the

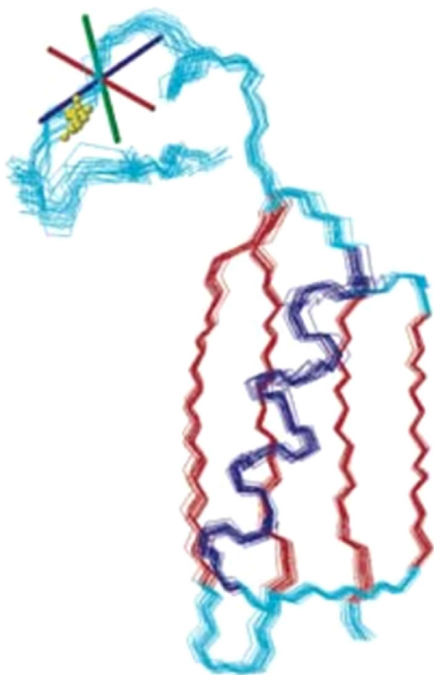


Fig. 5. Refined structure obtained using a two-point anchored LBP. The metal positions are indicated by yellow dots. The cross indicates the tensor orientation [59].

protein. Furthermore, labelling with an LBP leads to increased complexity of protein NMR spectra. In order to gain more versatility concerning the location of the attachment site, LBPs that are conjugated to cysteine residues were developed [116,117], but limited affinity for the lanthanides was often problematic.

In order to circumvent the drawbacks associated with LBPs, synthetic LCTs were developed. Based on chelators like diethylenetriaminepentaacetic acid (DTPA) [118,119] and ethylenediaminetetraacetic acid (EDTA) [21,120], synthetically produced metal binding sites with high lanthanide affinity can be obtained. These tags have to be attached to a unique ligation site on the protein. The thiol group of cysteine stands out as an ideal target due to its low abundance in protein sequences and the possibility for a chemoselective modification with LCTs. In order to provide a rigid tagging site in a well-defined region of the protein, cysteine residues in alpha-helices are typically chosen as attachment points. For the tagging of cysteine residues in proteins, suitable linkers had to be developed. The most practical systems are pyridine-2-thiol as well as methyl- and phenylsulphonyl activated linkers [1]. Unfortunately, maleimides that are frequently used in biochemistry are not suitable for the attachment of LCTs to proteins because they yield an additional stereo-centre at the linkage site upon tagging, resulting in two diastereomeric forms of the protein-tag construct. Recently, artificial amino acids, e.g., genetically encoded *p*-azido-*L*-phenylalanine residues, which can be tagged in a bio-orthogonal fashion, were introduced by Otting et al. [6]. Bio-orthogonal LCTs allow one to investigate proteins that cannot be expressed in a stable form with a single accessible cysteine residue. Furthermore, bio-orthogonal LCTs can be used to characterize proteins that bear multiple accessible cysteine residues that are essential to their biological function. Despite the fact that it requires considerable efforts to incorporate bio-orthogonal LCTs into biomacromolecules, this is definitively the most general approach and in many difficult cases of multi-cysteine proteins will be the only viable one.

2.2. Non-DOTA-based lanthanide chelating tags

Already in 2002, Dvoretzky et al. demonstrated the use of a pyridine-2-thiol activated linker moiety in combination with an EDTA-derived metal chelator for applications of paramagnetic NMR [121]. Upon ligation of the novel chelator, loaded with Yb^{3+} (among other paramagnetic metal ions) to barnase H102C by formation of a disulphide linkage, valuable PCS and RDC long-range structural restraints were analysed and fitted to the crystal structure of wild-type barnase.

The development of LCTs based on two-armed caged lanthanide NMR probes (CLaNP) was started in 2004, when Prudêncio et al. reported the first CLaNP, consisting of a DTPA moiety bearing two reactive groups suitable for ligation to cysteine residues [118]. Remarkably, CLaNP gave rise to observable PCSs detected on nuclei farther away than 40 Å from the metal centre. Unfortunately, up to five shifted resonances were detected for each residue. However, the strong PCSs induced by CLaNP clearly already illustrated at this point in time the potential of the use of PCSs and RDCs in structural biology.

In 2004, Ikegami et al. developed a novel, enantiomerically pure EDTA-derived LCT that can be attached via a cysteine residue to a protein of interest [21]. The LCT was then used for partial alignment of the protein in solution and RDCs with a magnitude up to 8 Hz were observed at a field strength of 18.8 T. Notably, the LCT was conjugated to a trigger factor, a ribosome-associated molecular chaperone, and an analysis of the tensors of the anisotropic magnetic susceptibility was performed. The study identified the drawback of EDTA-derived LCTs, i.e., the doubling of the number of peaks due to the generation of an additional stereo-centre on

the nitrogen atom close to the introduced linker, which becomes chiral upon metal complexation (Fig. 6).

Leonov et al. developed in 2005 an improved version of the EDTA-derived LCT presented by Ikegami et al. [21], giving rise to only one set of signals when attached to the protein of interest [62]. The diastereochemically pure compound can be synthesized in 5 steps and the corresponding (*R,R*)- and (*S,S*)-diastereomers were separated by high-performance liquid chromatography (HPLC). Since each of the two nitrogen atoms, which are able to coordinate to the lanthanide ion, are substituted with two identical groups, neither of them becomes stereogenic upon metal complexation and only one set of signals is detected, thereby significantly facilitating analysis of the acquired spectra.

Haberz et al. extended this work by the synthesis of two new EDTA chelators in 2006 that were successfully attached to the trigger factor protein and apo-calmodulin [122]. The authors could improve the precision of the solution structure of the trigger factor and showcase the extremely high affinity of the new EDTA-derived LCT for lanthanides, exceeding the one observed for the natural metal binding site of apo-calmodulin.

In 2008, Su et al. presented an LCT based on 4-mercaptomethyl-dipicolinic acid (4MMDPA) that can be readily attached to proteins via a disulphide linkage [10]. 4MMDPA loaded with Yb^{3+} yields PCSs of up to 2 ppm when attached to the cysteine-68 residue of arginine repressor from *E. coli*. Notably, the LCT is immobilized on the protein's surface by an additional coordinating carboxy group of the protein.

Man et al. developed in 2010 a small and rigid LCT, 3-mercapto-2,6-pyridinedicarboxylic acid (3MDPA) [123]. The applicability of this new LCT was illustrated by attachment to the N-terminal domain of the arginine repressor (ArgN) from *E. coli* and T4 lysozyme C54T C97A Q69C. The LCT yielded anisotropy parameters of $\Delta\chi_{\text{ax}} = 6.6 \times 10^{-32} \text{ m}^3$ and $\Delta\chi_{\text{rh}} = 2.4 \times 10^{-32} \text{ m}^3$ for thulium-loaded 3MDPA in the ArgN-3MDPA construct. While the simple synthesis and the sufficient resulting anisotropy parameters render the tag interesting for studies of protein structure and dynamics, possible interactions with the Glu21 sidechain of ArgN and the fact that only three donor atoms are used to coordinate the lanthanide

ion constitute a drawback of 3MDPA as LCT when compared to rigid and high affinity ligands.

In 2011, Swarbrick et al. developed a small LCT that can be attached to the α -helix of a protein of interest via a disulphide linkage (Fig. 7) and becomes rigidified on the protein by interaction of the lanthanide ion with an Asp residue in the $i + 4$ position [125]. While the synthesis of the presented iminodiacetic acid-tag (IDA-SH) is extremely simple and attractive (only three steps with an overall yield of 45%), the requirement of an Asp residue in the $i + 4$ position restricts the applicability of the LCT to proteins where a native cysteine (or one introduced by mutation) is located in an α -helix with an Asp residue in the $i + 4$ position. Swarbrick et al. further demonstrated that it is possible to use nitrilotriacetic acid-derived tags (NTA-SH) for the generation and analysis of large PCSs and RDCs [124]. Yagi et al. then investigated the optimal positioning of the IDA-SH and NTA-SH LCTs [126]. In a systematic study, in which the authors introduced the mentioned LCTs in an α -helix with either glutamate or aspartate residues in the positions $i - 4$ or $i + 4$, it was found that the largest anisotropy parameters can be observed for a NTA-SH LCT with a glutamate residue positioned at $i - 4$. Furthermore, it was shown that the NTA-derived LCTs are able to induce sufficiently large PCSs, independently of

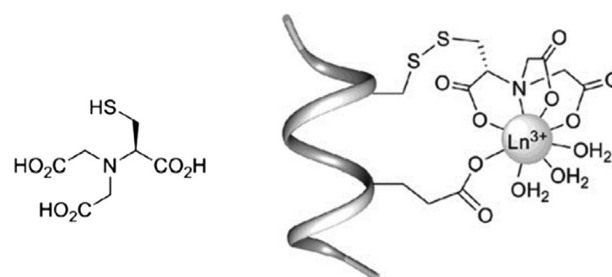


Fig. 7. Proposed lanthanide chelating mode of the NTA-derived LCTs in an α -helix with a carboxy sidechain in $i + 4$ or $i - 4$ position developed by Swarbrick et al. [124].

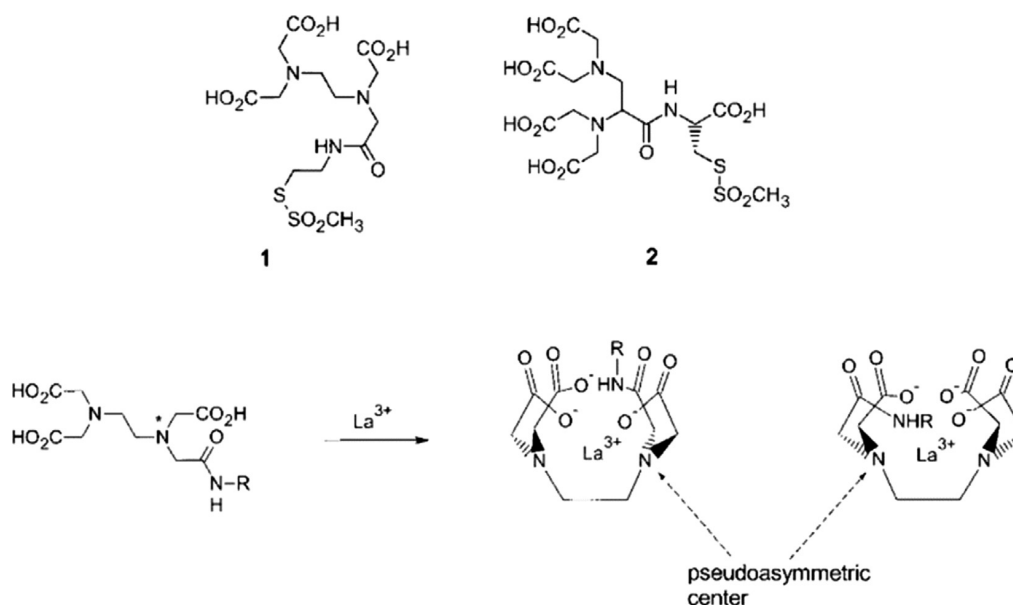


Fig. 6. Structure of EDTA-derived LCTs reported by Ikegami et al. [21] and Leonov et al. [62]. While compound 1 yields an additional stereo-centre upon metal complexation due to formation of a pseudoasymmetric centre, compound 2 is designed with two identical substituents on both nitrogen atoms coordinating to the lanthanide ion in order to avoid this behaviour.

nearby carboxylate groups in the protein. In contrast to the NTA-SH LCT, the IDA-SH LCT requires an aspartate residue in the $i + 4$ position.

Jia et al. presented in 2011 the use of the dipicolinic acid-derived tags like 4MDPA for the generation of structural restraints induced by the paramagnetism of lanthanides [127]. The switch from 3MDPA to 4MDPA is highly favourable in order to suppress rotational averaging of the paramagnetic effects, since the short and rigid tether, as well as the symmetric assembly of the 4MDPA chelator and the lanthanide, ensure a successful transfer of the paramagnetic effects from the lanthanide to the protein (Fig. 8). Further advantages are the small size of the tag that ensures little disturbance of the protein structure and the quantitative tagging yields. However, the anisotropy tensor parameters of $\Delta\chi_{ax} = -9.2 \times 10^{-32} \text{ m}^3$ and $\Delta\chi_{rh} = -5.4 \times 10^{-32} \text{ m}^3$ for Tm^{3+} on the ArgN construct and the possible interactions of DPA-derived LCTs with other residues of the protein [123] still offers room for further developments of LCTs.

In 2011, Peters et al. reported the synthesis of Cys-Ph-TAHA, a highly symmetric LCT based on the tris(2-aminoethyl)aminehexaacetic acid (TAHA) chelator [14] (Fig. 9). This LCT induces PCSs of up to 2 ppm and RDCs of up to 18 Hz, while providing a metal affinity in the femtomolar range and only a single set of signals in ^1H - ^{15}N HSQC spectra. Furthermore, the LCT shows a good thermal stability in the temperature range of 278–315 K. However, due to the carboxylic acid ligands, the LCT is pH-sensitive, so that the LCT cannot be used below pH 4.0, for example to monitor protein unfolding. In order to show that the LCT is suitable for analysis of large proteins and their complexes, Cys-Ph-TAHA was successfully attached to a

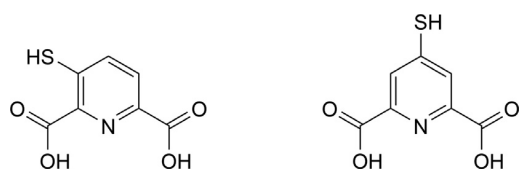


Fig. 8. Comparison of the structures of 3MDPA [123] (left) and 4MDPA [127] (right). The symmetry of 4MDPA reduces the number of possible alternative conformations that could result in decreased PCS [127].

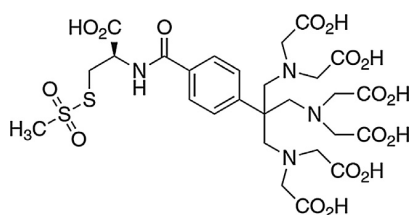


Fig. 9. Structure of the Cys-Ph-TAHA LCT [14].

90 kDa assembly of a Lac repressor protein in complex with DNA and isopropyl β -D-1-thiogalactopyranoside (IPTG).

An LCT ligated via a thiol-ene reaction to the protein was proposed in 2011 by Li et al. [128]. The LCT can be attached by incorporating a vinylpyridine moiety to ArgN and human ubiquitin. The protein-tag conjugate is stable in the presence of reducing agents such as dithiothreitol (DTT) or 3,3',3''-phosphanetriyltripropanoic acid (TCEP) due to the inert thioether linkage. Tagging was achieved in a quantitative manner using an eight-fold excess of the LCT. Low molecular weight compounds could subsequently be removed by ion-exchange chromatography leading to the desired tag-protein conjugate in 80% yield. Sizeable anisotropy parameters of $\Delta\chi_{ax} = 13.8 \times 10^{-32} \text{ m}^3$ and $\Delta\chi_{rh} = 7.5 \times 10^{-32} \text{ m}^3$ were induced on ubiquitin T22C using the Tm^{3+} -loaded LCT.

Yang et al. described in 2012 the synthesis of 4-vinyl(pyridine-2,6-diyl)bismethylenetriolo tetrakis(acetic acid) (4VPyMTA) and its use as chelator in paramagnetic NMR [129]. The presented tag shows only one set of signals, the formed linkage of the LCT to the protein is stable under reducing conditions and the affinity for lanthanides is higher than that of EDTA. The induced anisotropy is only in a medium range, i.e., $\Delta\chi_{ax} = -2.3 \times 10^{-32} \text{ m}^3$ and $\Delta\chi_{rh} = -0.5 \times 10^{-32} \text{ m}^3$ for Tm^{3+} on ubiquitin E64C, however, this offers two advantages: (i) convenient assignment of ^1H - ^{15}N HSQC spectra and (ii) a large number of PCS can be detected. The development of reduction-stable linker moieties is highly desirable, however, the ligation yield of 80% still leaves room for further improvement.

As a further development to the earlier introduced dipicolinic acid LCTs, Huang et al. reported the synthesis and application of a related 4'-mercapto-2,2': 6',2''-terpyridine-6,6''-dicarboxylic acid (4MTDA) LCT [130]. Upon attachment of the LCT to the protein of interest, interactions of a carboxylate sidechain of the protein with free coordination sites of the lanthanide in the chelator immobilize the LCT on the protein's surface. When attached to the protein, the presented LCT yields large PCSs and RDCs of up to 13 Hz. Furthermore, the described LCT can also be used for fluorescence and EPR studies. Unfortunately, 4MTDA exhibits a reactivity that is inferior to the related 4MMDPA [10] and 3MDPA [123], since the ligation yields are only 50–70%.

Yang et al. described two LCTs with a phenylsulfonyl leaving group (4-(phenylsulfonyl)-pyridine-2,6-dicarboxylic acid (4PS-DPA) and 4-phenylsulfonyl-(pyridin-2,6-diyl)bismethylenetriolo tetrakis(acetic acid) (4PS-PyMTA)) leading to an extremely short and rigid thioether linkage with the protein of interest [131]. Tm^{3+} loaded 4PS-PyMTA leads to large observed anisotropy parameters of $\Delta\chi_{ax} = 16.6 \times 10^{-32} \text{ m}^3$ and $\Delta\chi_{rh} = 3.9 \times 10^{-32} \text{ m}^3$ when attached to ubiquitin G47C R72A R74A. Due to the generation of reduction-stable protein-LCT linkages (Fig. 10), the authors anticipated that the phenylsulfonylpyridine moiety should find broad applications to structural biology and related fields.

In 2015, Loh et al. presented "clickable" IDA and NTA tags bearing an alkyne linker moiety that can be ligated by a copper catal-

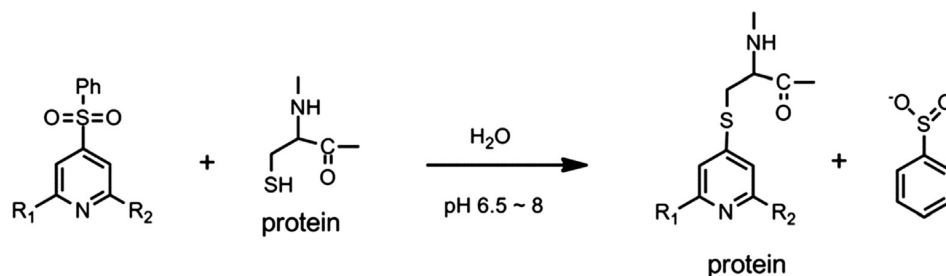


Fig. 10. Tagging reaction of the 4PS-DPA and 4PS-PyMTA LCTs bearing a phenylsulfonyl moiety reported by Yang et al. [131].

used azide alkyne click reaction to a site-specifically introduced *p*-azido-*L*-phenylalanine (AzF) residue in ubiquitin or in GB1; they further presented an analysis of paramagnetic effects and anisotropy parameters [132]. For the thulium-loaded alkyne-NTA LCT, maximal anisotropy parameters of $\Delta\chi_{ax} = 16.3 \times 10^{-32} \text{ m}^3$ and $\Delta\chi_{rh} = 7.8 \times 10^{-32} \text{ m}^3$ were observed upon ligation to the GB1 V21 construct. The authors suggest using 150 mM NaCl during the ligation reaction, which proved to be beneficial for the AzF mutants of GB1, ERp29, PpiB, the intracellular domain of the p75 neurotrophin receptor and the West Nile virus NS2B-NS3 protease in terms of precipitation of the protein during the attachment of the LCT.

In line with the findings of Loh et al., Jiang et al. developed diethylene-triamine-tetraacetate propyl-1-yne (DTTA-C3-yne) and diethylene-triamine-tetraacetate butyl-1-yne (DTTA-C4-yne) LCTs that were attached in bio-orthogonal fashion to ubiquitin or to the enzyme E II B, using a copper catalysed azide-alkyne click reaction [133]. Upon attachment to ubiquitin incorporating a *p*-azido-*L*-phenylalanine (AzF) residue in position 18, the Tb^{3+} and Tm^{3+} loaded LCTs yielded sizeable anisotropy parameters for the analysed PCSs and delivered RDCs of up to 8.3 Hz. Furthermore, the authors presented a refined model of the interactions in the UBA1-ubiquitin complex based on intermolecular PCSs.

Chen et al. reported in 2017 the use of DTPA-derived LCTs, i.e., 4PS-PyDTTA and its methyl-substituted analogue 4PS-6M-PyDTTA [134]. Both of the investigated LCTs yield large PCSs and RDCs, more specifically, the methyl-substituted LCT generally shows larger anisotropy parameters for ubiquitin or SrtA with the thulium-loaded LCT, most likely due to suppressed rotational averaging caused by the additional methyl group. The presented LCT offers a high stability of the formed complexes – neither 1.0 eq. of EDTA in the ubiquitin sample, nor the naturally present calcium binding motif in SrtA was able to interfere with the binding of the lanthanide ions by the DTPA chelator.

The developed EDTA-derived LCTs can lead to peak doubling observed for the protein in ^1H - ^{15}N HSQC experiments and therefore to two sets of signals with two associated χ -tensors, since some of the EDTA ligands become chiral upon metal complexation and lead to two diastereomeric species in the protein-tag conjugate. Furthermore, the obtained PCSs were only of modest size. Using DTPA-based tags, up to eight different sets of signals can be expected. For two-armed lanthanide probes, namely CLaNP, up to five different sets of signals were observed in practice [118] (Fig. 11). In order to overcome the shortcomings of EDTA, DTPA and related LCTs, several research groups developed DOTA-based LCTs, which often show superior performance.

2.3. DOTA-based lanthanide chelating tags

DOTA (1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid) provides an ideal framework for the chelation of lanthanide ions and development of LCTs. Besides the fact that DOTA complexes of lanthanide ions display a very high thermodynamic stability (equilibrium constant for the formation of $\text{Gd}(\text{III})$ -DOTA = 10^{25} M^{-1} at pH 7 [135]), they can also form neutral lanthanide complexes, when suitably modified. Due to the excellent properties of DOTA as ligand in Gd^{3+} complexes, this chelator has found widespread applications in MRI as well as in radiopharmacy [136,137].

The first DOTA-based LCT was reported by Häussinger et al. in a conference abstract in the year 2005 [138]. Based on the DOTA framework that has a binding constant that is 10 orders of magnitude larger than that of EDTA, a lanthanide tag equipped with a short $\text{SCH}_2\text{CH}_2\text{NH}$ linker between the cysteine sulphur of the protein and the carboxy group of the DOTA framework was developed and applied to characterize the 50 kDa dimer of the calcium dependent

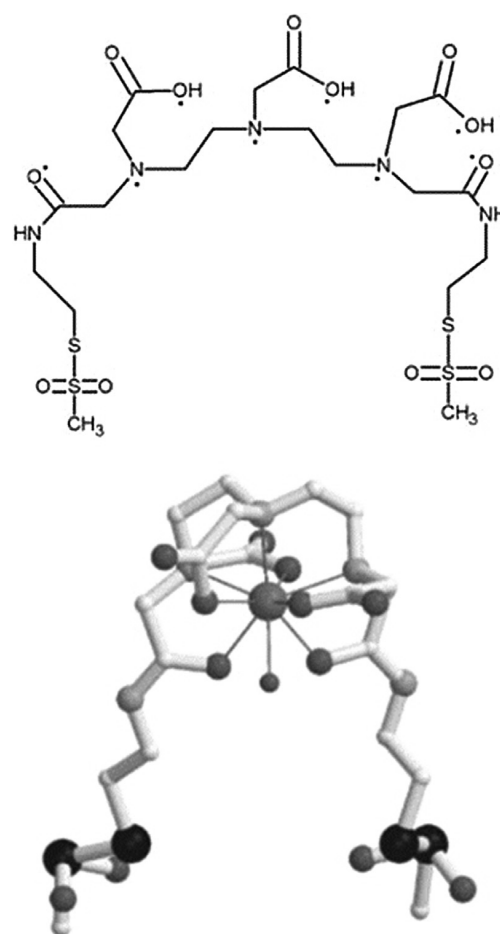


Fig. 11. Structure of the DTPA-based two-point anchored LCT CLaNP (dots: coordinating atoms) [118].

cell adhesion protein ECAD12. This DOTA-based LCT can be used over a wide pH range and in the presence of EDTA and phosphate buffer.

However, the complicated conformational mobility of the DOTA ligand, i.e., the two different conformations of the basic ring scaffold (denoted $\delta\delta\delta\delta$ and $\lambda\lambda\lambda\lambda$) and the two different conformations of the pendant acetate arms of the chelator (denoted Δ and Λ), in principle lead, upon attachment to the protein of interest, to four different conformational isomers ($\Delta(\delta\delta\delta\delta)$, $\Delta(\lambda\lambda\lambda\lambda)$, $\Lambda(\delta\delta\delta\delta)$, $\Lambda(\lambda\lambda\lambda\lambda)$) (Fig. 12). For certain residues in ubiquitin S57C labelled with the Dy-DOTA-SSPy LCT up to three different resonances were observed in ^1H - ^{15}N HSQC spectra. Furthermore, it is reasonable to suspect that the ring flexibility of Ln-DOTA-SSPy can lead to a significant motional averaging of the PCS in the case of fast conformational exchange.

In order to decrease the ring motions of the DOTA-derived LCT resulting in motional averaging of PCS, methyl substituents were introduced both on the side arms and the macrocyclic polyamine scaffold [9] (Fig. 13). The methyl group substituents on the nitrogen-containing macrocycle adopt an equatorial-upper position [119,139] when the ligand is coordinated to a lanthanide metal ion and prevent motional averaging and line broadening of signals by locking the 12-membered ring in a $(\delta\delta\delta\delta)$ conformation [9] (Fig. 14).

This strongly paramagnetic, extremely rigid LCT displayed very large PCSs of over 5 ppm and RDCs over 20 Hz when attached to ubiquitin S57C. Besides the fact that these properties were

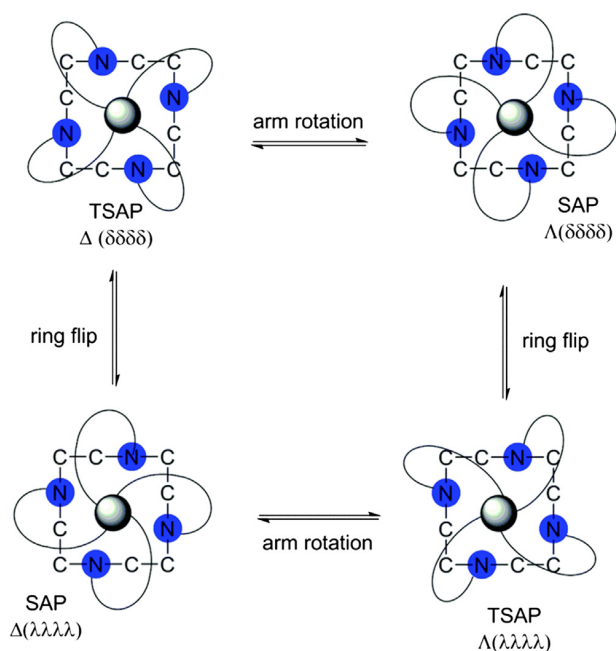


Fig. 12. Conformations of DOTA-derived ligands upon metal coordination [139].

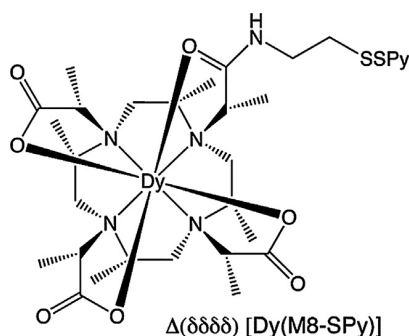


Fig. 13. Structure of the rigidified DOTA-based Dy-DOTA-M8-(8S)-SSPy in TSAP conformation [9]. (Reproduced with permission from Ref. [9]. Copyright 2009 American Chemical Society.)

unprecedented for a single-point attached LCT, the tag can be used under extreme buffer conditions, due to the high stability of the lanthanide complex over a wide pH range and very high affinity of the chelator towards lanthanide ions.

Interestingly, Dy-DOTA-M8-(8S)-SSPy exhibits two sets of signals when attached to a protein of interest resulting in a complicated analysis of ^1H - ^{15}N HSQC spectra containing the PCSs and RDCs and hence the structural restraints for the investigation of the structure of the protein. Opina et al. reported that the two sets of signals originate in the presence of two different conformational species of the LCT due to repulsive interactions of two methyl groups, such that the two conformers interconvert by rotation of the side arms of the Ln-DOTA-M8-(8S)-SSPy LCT on a timescale of approximately 4 h [139]. Opina et al. further showed that the two conformational species of the Ln-DOTA-M8-(8S)-SSPy LCTs can be observed in significant amounts for members of the lanthanide series between Eu^{3+} and Yb^{3+} and that the observed ratio of the two conformational species is governed by the ionic radius of the chelated lanthanide ion [139]. Strickland et al. investigated the temperature and pH-sensitivity of paramagnetic susceptibility

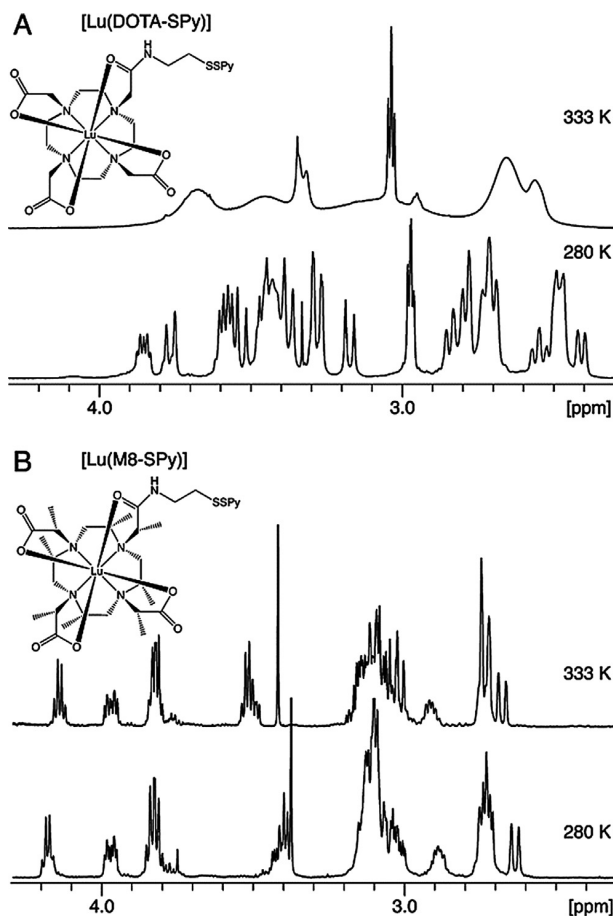


Fig. 14. (A) Motional averaging of NMR signals using a non-rigidifying DOTA chelator. At 280 K, broadening due to the motions is clearly decreased. (B) For the methyl-substituted chelators, even at elevated temperatures the motions are significantly decreased when compared to the non-methyl-substituted analogue. All spectra were acquired at 600 MHz proton frequency [9]. (Reproduced with permission from Ref. [9]. Copyright 2009 American Chemical Society.)

tensor of Ln-DOTA-M8-(8S)-SSPy LCTs [28]. The presented study investigating different lanthanide complexes showed that the temperature sensitivity depends on the size of the lanthanide ion and the conformational properties of the ligand. More specifically, a larger lanthanide ion such as Dy^{3+} leads to coordination of an apical water molecule and higher sensitivity towards temperature changes, while for smaller lanthanides such as Yb^{3+} , ^{17}O NMR experiments showed that no water molecule is coordinated to the LCT and hence the induced anisotropy does not significantly vary over a temperature range of 288–308 K [28]. The effect on the paramagnetic susceptibility tensor over a pH range of 5.0–7.4 was shown to be not significant [28]. Since the publication of the Dy-DOTA-M8-(8S)-SSPy LCT in 2009, it was applied to a range of projects. Among others, these include localization of inhibitors on proteins [71], exploration of the metal position of mobile spin labels with new fitting routines [140] as well as studies of extremely large protein complexes [12].

Joss et al. reported that the Ln-DOTA-M8-(4R4S)-SSPy stereoisomer of the Ln-DOTA-M8-(8S)-SSPy LCT shows only one conformational species for the whole lanthanide series [23]. Furthermore, the paramagnetic susceptibility of the lanthanide chelated by the ligand is significantly altered when compared to its predecessor Ln-DOTA-M8-(8S)-SSPy, a result that provides evidence for the influence of the coordination polyhedron and the ligand-metal

interactions on the magnetic anisotropy, as was proposed earlier by Mironov et al. [141].

Besides the inversion of the stereochemistry of the eight-fold methyl substituted LCT, introducing bulky and large groups on the side arms in a LCT can also lead to suppression of the conformational interchanges of the pendant arms and basic ring scaffold. Graham et al. presented a DOTA-amide LCT inducing large PCSs and exhibiting only one conformational species and hence only one set of signals [63]. The increased anisotropy transferred from the LCT to the protein when compared to other non-methyl substituted LCTs originates most likely in the strongly suppressed rotation of the LCT on the protein's surface due to the large and sterically demanding side arms, a result that is supported by the surprisingly small intrinsic shift range of 170.5 ppm in ^1H spectra of the ytterbium complex when compared to the anisotropy induced on the protein.

Lee et al. introduced a DOTA-derived, compact and hydrophilic LCT with a significantly shorter disulphide linker when compared to the DOTA-M8-(8S)-SSPy LCT that can be synthesized in 8% overall yield [25]. Although the LCT only includes methyl substituents on the pendant arms and none on the basic ring scaffold, this LCT leads to observation of large anisotropy parameters ($\Delta\chi_{\text{ax}} = 26.6 \times 10^{-32} \text{ m}^3$, $\Delta\chi_{\text{rh}} = 6.0 \times 10^{-32} \text{ m}^3$, for $\text{Ln} = \text{Dy}^{3+}$) on ubiquitin A28C. The results obtained by Lee et al. show in an impressive manner the large influence of the tagging site on the observed anisotropy parameters. For ubiquitin A28C anisotropy parameters of $\Delta\chi_{\text{ax}} = -19.4 \times 10^{-32} \text{ m}^3$ and $\Delta\chi_{\text{rh}} = -7.8 \times 10^{-32} \text{ m}^3$ for the thulium-loaded LCT are observed, while for the 20 kDa-sized protein construct HPPK S112C C80A anisotropy parameters of $\Delta\chi_{\text{ax}} = 54.5 \times 10^{-32} \text{ m}^3$ and $\Delta\chi_{\text{rh}} = 12.5 \times 10^{-32} \text{ m}^3$ were reported. Due to the large induced anisotropies and its convenient synthesis, this LCT is highly promising for use in a diverse range of applications.

Based on the chelators introduced by Lee et al. [25], the same group developed the corresponding series of two-armed LCTs [142]. As expected for the switch from a single-armed LCT to a two-armed LCT, the anisotropy parameters induced by the two-armed LCTs are significantly improved by a factor of ~ 2.5 (anisotropy parameters for ubiquitin E24C A28C: $\Delta\chi_{\text{ax}} = 47.1 \times 10^{-32} \text{ m}^3$ and $\Delta\chi_{\text{rh}} = 18.6 \times 10^{-32} \text{ m}^3$, for $\text{Ln} = \text{Tm}^{3+}$). PCSs induced by LCTs bearing hydroxyl donor groups were reported to depend on pH [25,143]. However, Lee et al. observed for the mentioned two-armed LCT only a minor pH-sensitivity.

In order to provide a reduction-stable linkage suitable for in-cell applications of the DOTA-M8 LCT, Müntener et al. used a pyridine derived linker that reacts with a cysteine residue of the protein to give a stable thioether linkage (Ln-DOTA-(3R4S)-M7Py) [3] (Fig. 15). In order to improve the tagging reaction, an additional fluorine substituent was introduced on the pyridine in the ortho position of the leaving group, thereby stabilizing the Meisenheimer complex formed upon reaction of the pyridine linker of the LCT with the cysteine [144]. The observed magnetic anisotropy parameters for Ln-DOTA-(3R4S)-M7Py are similar to Ln-DOTA-M8-(4R4S)-SSPy [23]. This result can be explained by an improvement of the paramagnetic effects due to the shorter and more rigid linker, but balanced by a simultaneous decrease due to the unfavourable rotation of the LCT on the surface of the protein caused by the less optimal linkage angle [144]. A similar approach was published by Pan et al. shortly after, illuminating the potential of this method for the investigation of protein structures in living cells [2].

In order to attach DOTA-derived LCTs to unnatural amino acids, Loh et al. presented two LCTs that can yield tag-protein conjugates by a copper catalysed azide-alkyne cycloaddition to a *p*-azido-*L*-phenylalanine residue [6]. The described LCTs induce reasonable PCS on ubiquitin Glu18AZf constructs and the bio-orthogonality of the linkage presents also an interesting approach for labelling

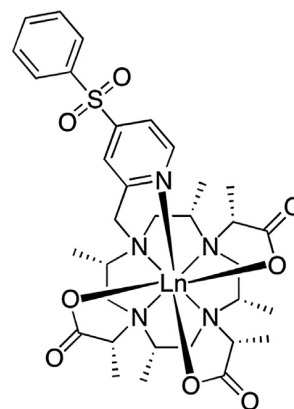


Fig. 15. Structure of Ln-DOTA-(3R4S)-M7Py suitable for in-cell NMR in oocytes reported by Müntener et al. [3]. (Reproduced with permission from Ref. [3]. Copyright 2016 American Chemical Society.)

of biomacromolecules other than proteins. However, the sample needs to tolerate copper ions, and in this connection it should be noted that His₆ tags inhibit the copper-alkyne click reaction and hence formation of the tag-protein linkage.

A further approach towards the applicability of the Ln-DOTA-M8 LCTs for in-cell NMR is provided by the introduction of a carbamidemethylene linker bearing an iodide leaving group on the acetyl amide moiety that is reactive in a selective fashion towards thiol nucleophiles [4]. Hikone et al. modified the Ln-DOTA-M8-(8S)-SSPy LCT accordingly (Ln-M8-CAM-1) and demonstrated the PCS analysis of M8-CAM-tagged ubiquitin mutants that were incorporated in HeLa cells by electroporation [4]. The stable thioether linkage formed upon tagging enables the use of the tag under reductive conditions. However, when compared to the original Ln-DOTA-M8-(8S)-SSPy LCT, the anisotropy is severely decreased due to the significantly longer, flexible linker part (Fig. 16).

As a next step in the development of the DOTA-M8 tags, an LCT with isopropyl substituents on the basic ring scaffold (Ln-P4M4-DOTA) was synthesized and tested with several protein mutants [145] (Fig. 17). Since the 1D spectrum shows a similar shift range as the one of DOTA-M8-(4R4S)-SSPy, the donor atoms were shown by DFT calculations to have a positional deviation of only 0.11 Å when compared to its eight-fold methyl-substituted predecessor, and the metal centre mobility was shown by the proposed methodology of Suturina et al. [140,146] to correspond to that of its predecessor, it can be concluded that the enhancement of the PCSs by a factor of almost 2 can most likely be attributed to a more hindered rotation of the LCT on the proteins surface caused by the isopropyl substituents on the basic ring scaffold.

Besides the investigation of isopropyl substituents by Joss et al., a new reductively stable pyridinethiazole linker moiety was proposed by Müntener et al. and attached to the heptamethyl-substituted LCT to obtain the Ln-M7PyThiazole-SO₂Me-DOTA LCT [144]. It was shown that the enhanced rigidity of the linker moiety (relative to the rather flexible LCT-CONH-CH₂-CH₂-S-Cys^{prot} linkage), the more favourable attachment angle (leading to decreased rotational averaging of the induced paramagnetic effects) and the improved synthetic access together provide an extremely favourable LCT in terms of synthetic access, stability of the tag-protein linkage under reductive conditions as well as magnitude of observed PCSs and RDCs [144].

An obvious extension was to test the newly developed linker moiety on an LCT equipped with an isopropyl-substituted backbone. The synthesized Ln-P4T-DOTA LCT shows large magnetic anisotropies induced by the different lanthanide complexes upon

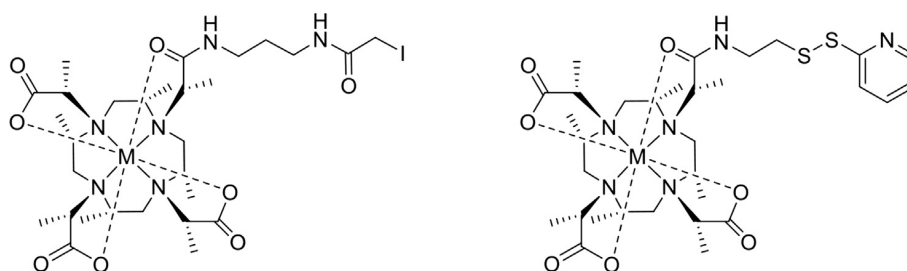


Fig. 16. Comparison of the structure of Ln-DOTA-M8-CAM-I reported by Hikone et al. [4] (left) and Ln-DOTA-M8-SSPy reported by Häussinger et al. [9] (right).

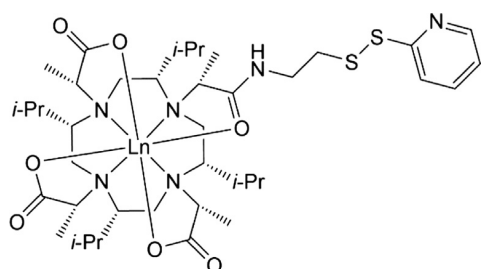


Fig. 17. Structure of a lanthanide complex of P4M4-DOTA in SAP conformation [145].

attachment to ubiquitin S57C, ubiquitin K48C and human carbonic anhydrase II (hCA II) S166C C206S and hence induces strong paramagnetic effects, i.e., both PCS and RDC, in the tag-protein conjugates [147]. Furthermore, these large paramagnetic effects could also be obtained under reductive conditions, e.g. to derive *de novo* derived in-cell protein structures from 2D NMR experiments [3].

The currently largest anisotropies induced by DOTA-derived LCTs are induced by a single-armed phenylsulphonated pyridine derivative of DOTA investigated by Yang et al. that constitutes a rigid and stable paramagnetic tag for analysis of protein structures and produces very large PCSs [148] (Fig. 18). However, no conversion was observed in the tagging reaction with pH 7.8 or below and the lanthanide complex could only be attached to the protein using harsh conditions for protein tagging, i.e., pH 9.0 over 16 h or pH 9.3 over 18 h, hampering the general applicability of this newly developed LCT. While the extremely large PCSs can result in a cumbersome assignment process of the paramagnetic signals for small- to medium-sized proteins, the presented LCT would be very suitable for applications on large-sized proteins as well as the localization of ligands within proteins over large distances.

Besides the introduction of more steric bulky substituents on the basic ring scaffold and more rigid linkers causing less motional averaging of the paramagnetic effects, the LCT presented by Keizers et al. (CLaNP-5) that bears two linker arms constitutes an interesting approach [24,149]. Although LCTs equipped with two attach-

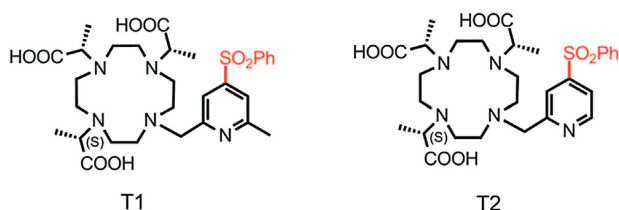


Fig. 18. Structures of the DOTA-derived LCTs reported by Yang et al. [148]. The additional methyl substituent in T1 leads to a decrease of motional averaging and thus rise to the extremely large PCSs observed.

ment sites to the protein require more challenging design and engineering of protein double mutants, which contain two cysteine residues separated by an optimal distance for the formation of the two disulphide linkages, the LCT presents an important advantage for application to the investigation of ligand-protein interactions when compared to single-armed LCTs, i.e., the induced paramagnetic effects are predictable and independent of the environment of the LCT on the protein surface. Such predictions can be used to generate a low-resolution model for the localization of a weakly binding ligand within a protein using PCS, removing the need for isotope labelling of both the ligand and the protein [69]. An initial low-resolution model of the position of the ligand can give important structural insights into protein-fragment hit complexes in fragment-based drug discovery (FBDD) studies. When one wishes to localize strongly binding ligands, ^{19}F PCS and a full assignment of ^1H - ^{15}N HSQC spectra of paramagnetic protein-tag-ligand conjugates can provide the position of the ligand with high accuracy [71].

The CLaNP-3 LCT reported in 2007 by Vlasie et al. [150] is an example of an earlier developed DOTA-derived CLaNP LCT. The two-armed LCT exhibits multiple isomers that are in exchange, but to the delight of the researchers only one signal was observed for the majority of the amides in ^1H - ^{15}N HSQC experiments, which paved the way for the determination of one single set of parameters of the induced anisotropy. By using the Yb^{3+} loaded LCT, PCSs were observed over a maximal distance of 35 Å away from the metal centre.

In order to reduce the rather high, three-fold positive charge of the CLaNP-5 LCTs that can possibly lead to spurious interactions with proteins, Liu et al. substituted the N-oxide moieties by nitro-substituted phenol groups, giving rise to an LCT containing two negatively charged alkoxides upon coordination of the three-fold positively charged metal centre, resulting in the singly positively charged CLaNP-7 LCT [152]. For the CLaNP-7 LCT, the orientation of the magnetic susceptibility tensor is pH-dependent when a histidine residue is located in the immediate vicinity of the binding site, a feature that could originate in interactions of the histidine residue with an apically coordinated water/hydroxyl molecule. In a further study, Liu et al. introduced the CLaNP-9 LCT that delivers large PCSs on a Paz-Cu^{II} E51C E54C construct and possesses an overall charge of +1, similarly to CLaNP-7 [151] (Fig. 19).

Lee et al. reported in 2016 an enantiomeric pair of pyridine-2-thiol activated, two-armed LCTs that gives rise to two different alignments of the magnetic susceptibility tensor [143]. For the thulium-loaded LCT, large RDCs of up to 24.8 Hz and large PCSs were observed on a ubiquitin E24C A28C double mutant. Interestingly, the observed PCSs and associated anisotropy parameters at pH 6.5 are 40% smaller than those obtained at a pH of 8.0. Since the two LCTs can be readily synthesized in parallel and two different magnetic susceptibility tensors can be obtained using the same tagging site, these LCTs can be conveniently used for studies of protein structure and ligand binding.

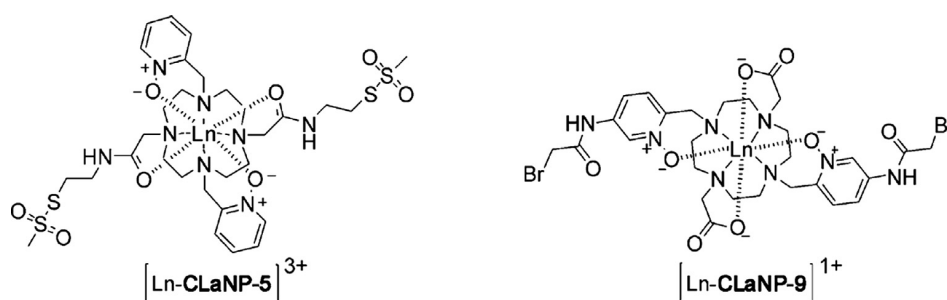


Fig. 19. Structural comparison of CLaNP-5 [24] and CLaNP-9 [151], which incorporates two carboxy groups that decrease the overall charge of the lanthanide complex.

Besides disulphide, pyridine-derived and iodoacetamide linkers, bromoacetamides were shown by Wu et al. to be of significant value for ligation of an LCT to a phosphorothioate group in DNA [30]. The reduction-stable linkage that is formed between the LCT equipped with a bromo leaving group and the phosphorothioate offered the possibility for the authors to detect PCSs on DNA and subsequently fit the corresponding anisotropic $\Delta\chi$ -parameters. DOTA-derived LCTs are ideal scaffolds for the development of LCTs that can be attached to nucleic acids, since the high thermodynamic stability and kinetic inertness of the synthesized complexes ensures that no interferences between the metal ion and the phosphate backbone of the nucleic acid occur. The stereogenic phosphorus atom generated by the introduction of a phosphorothioate group leads to different sets of observed PCS. However, it is possible to separate the diastereomers by reverse-phase HPLC on a semipreparative C18 column leading to only one set of signals observed [30], thus opening the avenue of the transfer of the PCS methodology from proteins to DNA and RNA.

2.4. Tabular overview of DOTA-derived LCTs

In the following, all DOTA-derived LCTs published to date are listed with their corresponding reported anisotropy parameters and the tagging conditions (Table 1). It is clear that the reported $\Delta\chi$ -values suffer from a large variability, as they are usually obtained by the single-point approximation, which strictly does not hold true in any case. In particular, pronounced mobility can lead to over-estimated $\Delta\chi$ -values in combination with elongated metal – attachment site distances [140,146,153]. But nevertheless the $\Delta\chi$ -values, especially when obtained for several different proteins, allow an estimation of the performance of a LCT – when taken with due precautions. It is noteworthy, that the range of anisotropy for a given LCT within the lanthanide metal series varies dramatically with minute differences of the coordination polyhedron geometry and donor atom set – in many cases Dy causes larger anisotropy than, say, Tm, but there are also several examples where the order is reversed.

2.5. Parameters in the design of LCTs

A unique feature of the structural restraints obtained by attachment of an LCT to a biomacromolecule when compared to NOESY experiments is the longer distance range that can be covered. The largest anisotropy parameters reported by Yang et al. for the Ln-DO3MA-6MePy LCT [148], i.e., $\Delta\chi_{ax} = -84.3 \times 10^{-32} \text{ m}^3$ and $\Delta\chi_{rh} = -17.2 \times 10^{-32} \text{ m}^3$ for a dysprosium-loaded LCT attached to an ubiquitin G47C protein construct, yield a distance limit of 207 Å assuming a threshold of significant chemical shift differences of 0.005 ppm. Based on the intrinsic tensor anisotropies currently investigated in our research group and a full transfer of the intrinsic anisotropy caused by the lanthanide to the protein by an

ideal, rigid attachment of the LCT to the protein under investigation, it is not unrealistic to monitor distances over 250 Å in the future, which would greatly extend the versatility of this technique for biological applications.

The accuracy and precision in the localization of ligands within biomacromolecules that can be reached using LCTs is in the best cases comparable to the resolution of X-ray structures and exceeds the currently achieved resolution for cryogenic electron microscopy (cryo-EM) [155]. As reported by Zimmermann et al., the localization of fluorinated inhibitors within human carbonic anhydrase was achieved accurately to within 0.8 Å [71]. When using LCTs for the localization of ligands within biomacromolecules, paramagnetic susceptibility tensors that intersect in a close to orthogonal fashion at the ligand position deliver the best results in terms of accuracy. To judge whether the isosurfaces of the tensors intersect in a favourable fashion, an angle score was introduced by Zimmermann et al. that allows one to give a quantitative estimate of the accuracy of the ligand localization using a given set of tensors [71]. Double-armed LCTs that yield predictable tensor orientations and magnitudes [69] are particularly suited for this approach. The PCS based method opens a route to highly accurate localization of ligands within biomacromolecules even for cases that are difficult to investigate using X-ray crystallography.

For even further improvement of LCTs for biomacromolecules, three important questions in the design of LCTs emerge: (i) “What kind of ligand delivers the best framework for the synthesis of LCTs in terms of stability and inertness of the lanthanide complexes?” (ii) “How is it possible to maximize the intrinsic anisotropy of the lanthanide complexes?”, and (iii) “How is it possible to transfer the full anisotropy of the lanthanide complex to the investigated biomacromolecule?”

Using a rigid framework like in DOTA-derived chelators, the lanthanide ion can be conveniently accommodated in the cavity and the tightly chelating ligand ensures chemical stability and inertness towards the environment of the LCT, which is a unique feature that clearly improves on the stability of open-chain chelator-based LCTs [9].

The investigation of the intrinsic anisotropies of lanthanide complexes in combination with their structural properties is subject of much current research [52,53,156,157]. Furthermore, Mironov et al. reported quantitative relations between ligand atom positions and geometries and the observed anisotropies [141]. However, rational design of complexes with enhanced paramagnetic properties is cumbersome, since it is not straightforward to estimate the position of the ligand atoms in an LCT that is envisioned to be synthesized. A more convenient approach towards obtaining large intrinsic anisotropy parameters is provided by the introduction of sterically demanding substituents on both the basic cyclen ring moiety as well as on the side-arms of the DOTA chelator. Rigidified DOTA-complexes provide shift ranges in one-

Table 1
Overview on DOTA-derived LCTs including tagged protein constructs, PCS- and RDC-derived observed anisotropy parameters, Ln-C_β distance as well as the used conditions for the tagging reaction.

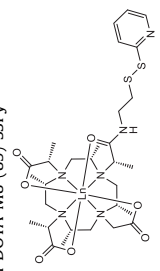
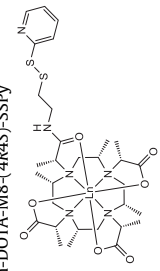
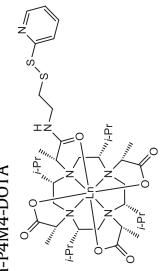
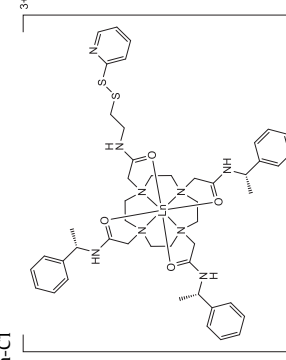
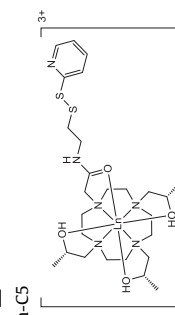
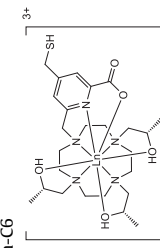
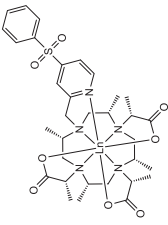
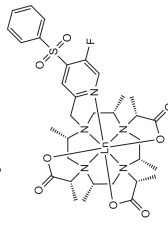
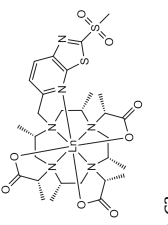
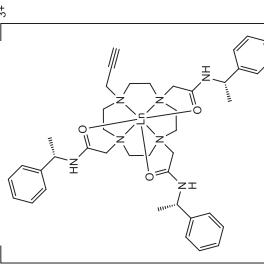
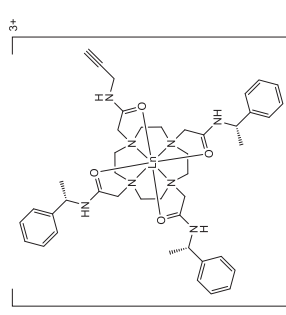
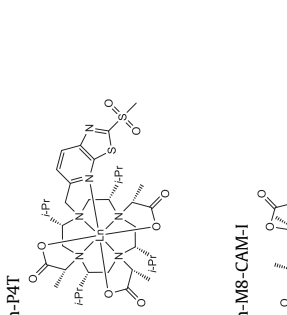
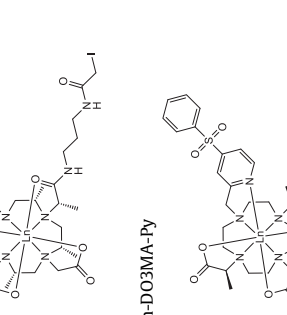
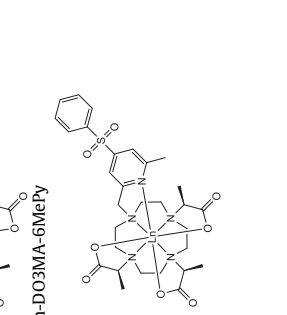

Name & structure	Protein construct	Ln ³⁺	$\Delta\chi_{ax, PCS}$ 10 ⁻³² m ³	$\Delta\chi_{rh, PCS}$ 10 ⁻³² m ³	$\Delta\chi_{ax, RDC}$ 10 ⁻³² m ³	$\Delta\chi_{rh, RDC}$ 10 ⁻³² m ³	Distance Ln-C _β [Å]	Tagging conditions & remarks	Ref.
Ln-DOTA-M8-(8S)-SSPy 	Ubi S57C (PDB 1D3Z)	Dy ³⁺	28	17	21	6	7.9	3.0 eq. LCT, pH 5.0, 10 mM phosphate, rt, 3 h	[9]
Ln-DOTA-M8-(4R4S)-SSPy 	Ubi S57C (PDB 1UBI)	Dy ³⁺ Tm ³⁺ Tb ³⁺	-8.7 19.6 -8.2	-4.1 3.0 -3.3	-	-	7.2	3.0 eq. LCT, pH 7.0, 10 mM phosphate, rt, 3 h	[23]
Ln-P4M4-DOTA 	hCA II S50C C206S (PDB 3KS3)	Dy ³⁺ Tm ³⁺ Tb ³⁺	-21.0 36.1 -16.7	-11.1 11.5 -11.8	-	-	7.6	3.0 eq. LCT, pH 6.8, 10 mM phosphate, rt, 3 h	[145]
	hCA II S166C C206S (PDB 3KS3)	Dy ³⁺ Tm ³⁺	-21.2 29.7	-4.5 7.6	-	-	6.6 5.9	6.0 eq. LCT, pH 6.8, 10 mM phosphate, rt, overnight	[145]
	Ubi S57C (PDB 1UBI)	Dy ³⁺ Tm ³⁺	-16.7 26.0	-7.5 1.0	-4.8 7.3	-	8.7 7.0	6.0 eq. LCT, pH 7.0, 10 mM phosphate, rt, overnight	[145]
	hCA II S166C C206S (PDB 3KS3)	Dy ³⁺ Tm ³⁺	-29.7 41.9	-11.5 17.1	-	-11.1	7.0 7.2	6.0 eq. LCT, pH 6.8, 10 mM phosphate, rt, overnight	[145]
Ln-C1 	ArgN C68 (PDB 1A0Y)	Dy ³⁺ Tb ³⁺ Tm ³⁺ Yb ³⁺	-29 -27 37 13	-11 -4 12 3	-	-	8.7	3.0 eq. LCT, pH 7.5, 50 mM HEPES, rt, 5 h	[63]
	Ubi A28C (PDB 1D3Z)	Dy ³⁺ Tb ³⁺ Tm ³⁺ Yb ³⁺	-10.8 -9.0 7.1 2.4	-3.5 -2.3 1.6 0.5	-	-	5.8	15.0 eq. LCT, pH 7.5-50 mM HEPES, rt, 3 h, then 3 °C, overnight	[25]
Ln-C5 	Ubi A28C (PDB 2MJB)	Dy ³⁺ Tb ³⁺ Tm ³⁺ Yb ³⁺	8.2 9.4 -18.7 -6.7	5.3 2.2 -6.9 -2.1	-	-	7.5 8.2 9.7 8.9	5.0 eq. LCT, pH 8.0, 50 mM HEPES, rt, 2 h	[25]
Ln-C6 	Ubi A28C (PDB 2MJB)	Dy ³⁺ Tb ³⁺ Tm ³⁺ Yb ³⁺	-9.4 -14.6 11.5 2.0	-5.9 -3.6 4.1 0.9	-	-	10.3 9.0 9.0 10.3	(1) 10.0 eq. Ellman's reagent, pH 8.0, (2) 50 mM HEPES, rt, 1 h (3) 3) 5.0 eq. LCT, pH 8.0, 50 mM HEPES, rt, 2 h	[25]

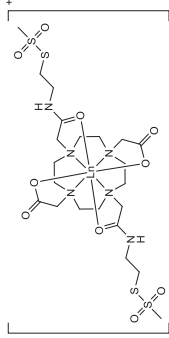
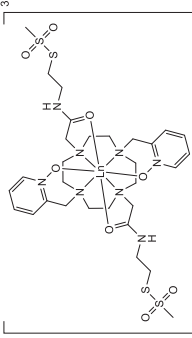
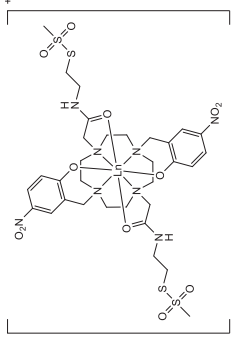
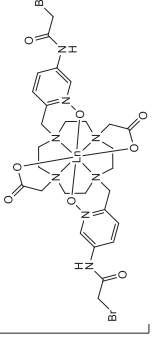
Table 1 (continued)

Name & structure	Protein construct	Ln ³⁺	$\Delta\chi_{\text{ax, PCS}}$ 10 ⁻³² m ³	$\Delta\chi_{\text{rh, PCS}}$ 10 ⁻³² m ³	$\Delta\chi_{\text{ax, RDC}}$ 10 ⁻³² m ³	$\Delta\chi_{\text{rh, RDC}}$ 10 ⁻³² m ³	Distance Ln-C _β [Å]	Tagging conditions & remarks	Ref.
Ln-M7Py 	GB1 E19C (PDB 2QMT) GB1 K28C (PDB 2QMT) GB1 E42C (PDB 2QMT)	Tm ³⁺ Tm ³⁺ Tm ³⁺	33.2 -32.0 -33.7	12.3 -18.2 -21.9	- - -	- - -	7.9 7.7 6.7	4.0 eq. LCT, pH 7.6, 50 mM phosphate, 0.5 mM TCEP, 40 °C, 24 h stable under reductive conditions	[3]
Ln-M7FPy 	Ubi S57C (PDB 1UBI) hCA II S166C C206S (PDB 3KS3)	Tm ³⁺ Tm ³⁺	17.8 -18.4	9.8 -4.4	- -	- -	6.6 5.4	6.0 eq. LCT, pH 7.0, 10 mM phosphate, 0.45 mM TCEP, rt, 16 h stable under reductive conditions	[144]
Ln-M7PyThiazole 	Ubi S57C (PDB 1UBI) Ubi K48C (PDB 1UBI) hCA II S166C C206S (PDB 3KS3)	Tm ³⁺ Dy ³⁺ Tb ³⁺ Tm ³⁺ Dy ³⁺ Tb ³⁺ Dy ³⁺	35.4 41.9 30.8 47.4 58.0 36.0 44.0	11.3 26.1 19.6 8.3 24.6 21.8 24.7	25.7 -36.3 -18.6 46.5 62.8 -39.8 42.9	3.8 -20.0 -8.3 12.8 34.4 -24.3 22.9	5.9 7.4 7.1	4.0 eq. LCT, pH 7.0, 10 mM phosphate, 0.1 mM TCEP, rt, 15 min stable under reductive conditions	
Ln-C3 	Ubi G18A2F (PDB 1D3Z)	Tb ³⁺ Tm ³⁺	9.0 -5.4	2.3 -1.7	- -	- -	9.1	10.0 eq. LCT, pH 7.5, 50 mM HEPES, 0.2 mM CuSO ₄ , 1.0 mM BITAA, 5.0 mM sodium ascorbate, rt, 16 h stable under reductive conditions	[6]

Ln-C4	Ln-P4T	Ln-M8-CAM-I	Ln-D03MA-Py	Ln-D03MA-6MePy
				
Ubi G18Azf (PDB 1D3Z)	Ubi S57C (PDB 1UBI) Ubi K48C (PDB 1UBI) hCA II S166C C206S (PDB 3KS3)	Ubi K6C (PDB 1D3Z) Ubi S57C (PDB 1D3Z)	Ubi G47C (PDB 1UBI) GB1 T11C (PDB 2QMT)	Ubi G47C (PDB 1UBI) GB1 T11C (PDB 2QMT)
Tb ³⁺ Tm ³⁺	Dy ³⁺ Tm ³⁺ Dy ³⁺ Tm ³⁺ Dy ³⁺ Tm ³⁺	Dy ³⁺ Dy ³⁺	Tm ³⁺ Tm ³⁺	Dy ³⁺ Tm ³⁺ Yb ³⁺ Dy ³⁺ Tm ³⁺
97 -9.7	54.0 39.3 -53.7 39.6 -46.1 44.3	23 6.8	22.2 21.8	-84.3 65.2 11.6 51.2 22.1
3.4 -3.0	27.0 14.6 -23.5 13.4 -30.5 4.8	12 2.4	13.00 13.1	-17.2 31.8 2.5 20.3 12.1
-	-45.0 -36.5 -	-	-	-
-	-21.8 -11.5 -	-	-	-
8.9	6.1 6.7 7.0	9.9 8.0	8.1 (to C _A) 8.8	8.9 (to C _A) 6.3
	6.0 eq. LCT, pH 7.0, 10 mM phosphate, 0.3 mM TCEP, rt, overnight stable under reductive conditions 10.0 eq. LCT, pH 6.8, 10 mM phosphate, 0.3 mM TCEP, rt, overnight stable under reductive conditions	2.0 eq. LCT, pH 7.4, 10 mM phosphate, 140 mM NaCl, rt, 1 week or 10.0 eq. LCT, pH 7.4, 10 mM phosphate, 140 mM NaCl, rt, 24 h stable under reductive conditions	3.0 eq. LCT, pH 9.0–9.5, 20 mM Tris, 0.2 mM TCEP, rt, 20 h stable under reductive conditions	
	[147]	[4]	[148]	

(continued on next page)

Table 1 (continued)

Name & structure	Protein construct	Ln ³⁺	$\Delta\chi_{\text{ax, RCS}}$ 10 ⁻³² m ³	$\Delta\chi_{\text{rh, RCS}}$ 10 ⁻³² m ³	$\Delta\chi_{\text{ax, RDC}}$ 10 ⁻³² m ³	$\Delta\chi_{\text{rh, RDC}}$ 10 ⁻³² m ³	Distance Ln-C _β [Å]	Tagging conditions & remarks	Ref.
	Paz E51C E54C (PDB 1PY0)	Yb ³⁺	5.9	1.9	-	-	10.0/9.0 (PRE optimized)	5.0 eq. LCT, pH 7.0, 20 mM phosphate, 4 °C, overnight	[150]
	Paz L8C K10C (PDB 1PY0)	Yb ³⁺	9.0	3.3	7.5	2.3	graphically displayed	10.0 eq. LCT, pH 7.0, 20 mM phosphate, 150 mM NaCl, 0 °C, 1 h	[24]
	Paz E51C E54C (PDB 1PY0)	Yb ³⁺	9.0	2.3	8.4	2.0			
	Paz E51C E54C (PDB 1PY0)	Eu ³⁺	1.5	0.1	-1.4	-0.7			
	Paz D100C S104C (PDB 1PY0)	Tm ³⁺	55.3	6.9	44.4	6.9			
	Paz D100C S104C (PDB 1PY0)	Dy ³⁺	-36.2	-15.0	-34.6	-11.0			
	Paz D100C S104C (PDB 1PY0)	Yb ³⁺	9.3	3.0	5.8	1.7			
	Cyt c N56C L58C (PDB 1YCC)	Yb ³⁺	8.1	2.5	-	-	7.3/7.7		[154]
	Cyt c N56C L58C (PDB 1YCC)	Tm ³⁺	50.5	6.0	-	-			
	Paz E51C E54C (PDB 1PY0)	Tm ³⁺	41.4	9.6	-	-		5.0 eq. LCT, pH 7.0, 20 mM phosphate, 150 mM NaCl, 4 °C, overnight	[152]
	T4 Lys N55C V57C (PDB 3DKE)	Yb ³⁺	4.2	5.6	-	-	graphically displayed		
	T4 Lys N55C V57C (PDB 3DKE)	Yb ³⁺	6.3	8.0	-	-			
	Paz E51C E54C (PDB 1PY0)	Yb ³⁺	7.5	2.3	-	-		10.0 eq. LCT, pH 8.0, 20 mM Tris, 150 mM NaCl, rt, 6 h	[151]
	T4 Lys N55C V57C (PDB 3DKE)	Tm ³⁺	39.5	10.6	-	-		stable under reductive conditions	
	T4 Lys N55C V57C (PDB 3DKE)	Yb ³⁺	6.3	2.2	-	-			
	T4 Lys N55C V57C (PDB 3DKE)	Tm ³⁺	32.5	15.6	-	-			

dimensional ^1H spectra of over 1000 ppm for their Tm^{3+} and Dy^{3+} complexes, thereby indicating the large intrinsic anisotropy of the lanthanide complex [23].

In order to transfer this large intrinsic anisotropy to the protein, motional and rotational averaging of the LCT on the protein has to be suppressed or at least substantially reduced. This was successfully achieved using two-armed LCTs [142,143] as well as by using large substituents on the basic cyclen ring [145] or on the side-arms of the chelator [63]. In the latter two examples, the large substituents effectively decrease the rotational averaging on the protein's surface. With the data available so far it is, however, not entirely clear whether substituted cyclen rings like tetramethylcyclen lead to improved anisotropy parameters mainly due to suppression of the $\delta\delta\delta - \lambda\lambda\lambda$ interconversion or due to more efficient immobilization on the protein surface – it is probable that both effects play vital roles.

The introduction of a suitable linker moiety is crucial for the successful design of an LCT. The important factors for success that have to be addressed in the design of LCTs are: (i) A short and rigid tether between the LCT and the protein should be achieved by the tagging reaction, (ii) The chosen activator should react in a fast and irreversible fashion, and (iii) The resulting linking site should be reduction-stable to enable NMR studies under reductive conditions, e.g. for in-cell applications.

The LCTs presented by Lee et al. combine a very short linker with the highly reactive pyridine-2-thiol activator, that is commonly used in PCS NMR spectroscopy [25,143]. In terms of providing a rigid linker moiety that enables fast tagging reactions resulting in a reduction-stable linkage, the pyridine-thiazolo linker carrying a methyl-sulphone as leaving group reported by Müntener et al. [144] displays favourable properties: (i) Ubiquitin S57C is tagged within 15 min. in an extent of >95%, (ii) The tagging reaction as well as the NMR experiments can be performed under reductive conditions, and (iii) The rigid linker moiety ensures a successful transfer of the anisotropy from the LCT to the protein resulting in large observed anisotropy parameters.

Although the limits in the design of LCTs have already been pushed very far, in the near future the combination of the approaches described in this chapter and the invention of novel linker moieties that can suppress the remaining motional averaging could lead to a significant boost in the observed anisotropy parameters and thus to an even higher distance range covered by LCTs and a larger accuracy in the localization of ligands within biomacromolecules due to a larger magnitude of the observed paramagnetic effects.

3. Applications of lanthanide chelating tags to the structural characterization of biomacromolecules

3.1. Applications of lanthanide chelating tags to proteins I: pseudocontact shifts

Since induced paramagnetic effects arising from LCTs, e.g. PCSs and RDCs, provide three-dimensional structural information, their use is highly valuable in the structural characterization of targets that are cumbersome to be analysed by other methods such as cryo-EM, X-ray crystallography and non-paramagnetic NMR. In a diverse range of cases, the data provided by other methods gives a valuable starting point for the structural characterization of proteins, since it allows one to start, for example, from a solid state structure that can be subsequently refined to a solution state structure, or by identifying flexible parts of the protein in solution.

Keizers et al. used a CLaNP-5 LCT loaded with Eu^{3+} , Gd^{3+} , Lu^{3+} and Tm^{3+} metal ions for the structural investigation of a protein complex of adrenodoxin and adrenodoxin reductase with a molecular weight of 65 kDa [158]. The observed PCSs were used as structural restraints in docking studies, in order to evaluate the binding between the two proteins in solution. The resulting structure closely matched the crystal structure of the investigated complex and the presented method holds great promise for the investigation of structures of protein complexes even under dilute conditions, since for the observation of the required PCSs only ^1H - ^{15}N HSQC spectra need to be acquired.

In 2011, Cruz et al. investigated conformational changes in the Dengue virus NS2B-NS3 protease using PCSs as structural restraints [159]. The protease was modified with multiple LCTs on several different tagging sites and the experimental PCSs were then used for the calculation of the structure of the Dengue NS2B-NS3 protease based on the crystal structure of the related West Nile virus protease. The results show that for positively charged, low-molecular weight inhibitors, the Dengue NS2B-NS3 protease adopts a closed state resembling that reported for the West Nile virus protease (Fig. 20).

In order to obtain a detailed model of the catalytic domain of the matrix metalloproteinase MMP-1, Bertini et al. attached the CLaNP-5 LCT containing Lu^{3+} , Yb^{3+} , Tm^{3+} and Tb^{3+} to the H132C K136C mutant of CAT MMP-1 [160]. Importantly, Bertini et al. showed, based on the observed PCS and RDC data, that the solution structure of CAT MMP-1 shows local rearrangements as well as significant secondary structure and domain rearrangements in solution relative to the previously obtained crystal structure.

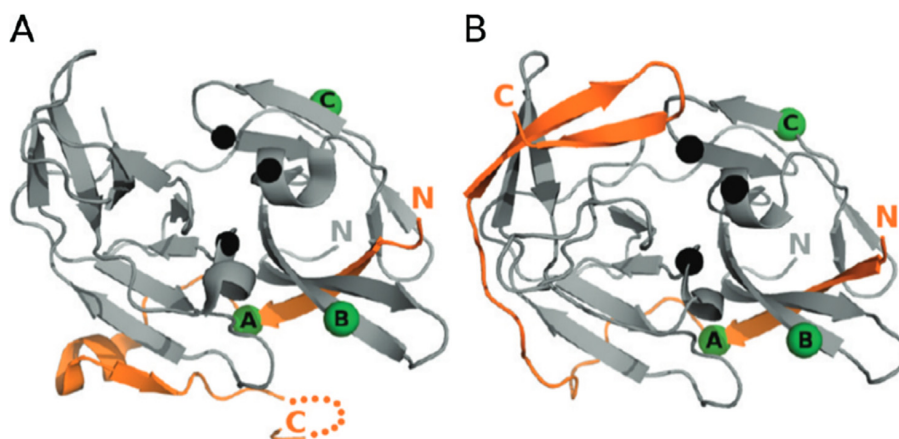


Fig. 20. Open (A) and closed (B) conformation of the C-terminal fragment of the dengue NS2B-NS3 protease investigated by Cruz et al. [159]. (Reproduced with permission from Ref. [159]. Copyright 2011 American Chemical Society.)

In 2013, it was shown by Li et al. that observation of PCSs in combination with Rosetta structural predictions can yield 3D structures of proteins using a magic angle spinning solid-state NMR approach [161]. GB1 was tagged with 4MMDPA, loaded, among other paramagnetic metal ions, with Yb^{3+} and Tm^{3+} . The introduction of the LCT gave rise to long-range structural restraints and the obtained structures using the observed PCSs in combination with the Rosetta structure calculations yield 3D structures similar to the X-ray structure model.

Yagi et al. attached two different LCTs (C1 and IDA-SH) to the chaperone Erp29, which is known to exhibit significant differences in the C-terminal domain between its solid state and solution state structures, and observed PCSs from four different tagging sites [162]. Based on the observed PCS data, the authors showed that the C-terminal domain of rat Erp29 in solution strongly resembles the reported crystal structure of the C-terminal domain of human Erp29. Since only ^1H - ^{15}N HSQC experiments have to be acquired for the structure refinement, the approach could be applied to more diluted protein samples. Furthermore, the presented approach is generally applicable for protein scaffolds, where the required tagging sites can be introduced by mutations.

In order to get structural insights into the complex of putidaredoxin and P450cam that catalyses the hydroxylation of camphor using the transfer of two electrons, Hiruma et al. determined both the solution state structure by paramagnetic NMR as well as the crystal structure by X-ray crystallography [163]. The researchers observed 446 structural restraints upon tagging of the protein complex using the CLaNP-7 loaded with Yb^{3+} and obtained an ensemble of 10 structures with an excellent RMSD of 1.3 Å. Furthermore, the authors determined the crystal structure that strongly resembles the 10 lowest-energy structures in solution.

In 2014, Chen et al. showed that the solution structure of the C-terminal β -hairpin of NS2B in the dengue virus NS2B-NS3 protease is different from the observed conformation in the crystal structure [164]. Importantly, the C-terminal segment, which was earlier thought to be located in an open conformation, was shown by analysis of PCSs to adopt a different structure, in which it is positioned close to the ligand binding site. The results provide important structural information for the design of new drugs against dengue virus.

The single-stranded DNA binding protein (SSBP) has two significant roles: on the one hand it protects single-stranded DNA and on the other hand it recruits single-stranded DNA processing enzymes. Although the unstructured C-terminal peptide segment of SSBP was shown to interact with the single-stranded DNA binding domain (OB-domain) by biochemical experiments with C-terminally truncated SSBP, no evidence for competitive binding of the C-terminal peptide fragment and a single-stranded DNA molecule was found in X-ray structures. Shishmarev et al. investi-

gated the protein using NMR spectroscopy and showed, based on NOESY data as well as PCSs induced by a lanthanide-loaded C1 tag [63] (Fig. 21), that the C-terminal fragment indeed binds to the OB-domain, suggesting direct competition between the C-terminal peptide fragment and single-stranded DNA in binding to the OB-domain [165].

Inspired by the results of Chen et al. [164], de la Cruz et al. tagged a two-component dengue virus NS2B-NS3 protease with two different paramagnetic metal tags [166]. In order to validate earlier results from X-ray crystallography concerning the arrangement of NS3 protease and NS2B, NS3 protease was modified with paramagnetic LCTs and PCSs were analysed to yield a magnetic susceptibility tensor. The results clearly show that the reported structures depict the interaction between NS2B and NS3 protease in a correct fashion. Furthermore, the authors showed that NMR signals that were believed to originate from the open state conformation, in fact are artefacts from protein degradation.

Camacho-Zarco et al. presented in 2014 the use of a reporter protein tagged with a Tm^{3+} containing CLaNP-5 LCT [167]. The paramagnetically labelled PDZ mutants can be fused to another protein of interest containing a C-terminal TGWETWV peptide sequence, giving rise to observation of paramagnetic effects on the unlabelled protein.

While the sparse nuclear Overhauser effect (NOE) signals failed to provide a 3D structure determination of the 7TM α -helical microbial receptor pSRII, Crick et al. showed that introduction of a paramagnetic metal tag yields PCS restraints that allow one to determine the correct three-dimensional fold of the membrane protein [168]. Since the small number of NOE signals available in this case typifies a problem that is often encountered in membrane proteins incorporating α -helical structures, the authors suggest that the acquisition and analysis of PCSs may be a general method to overcome this issue.

Analysis of PCSs by Brewer et al. solved a long-standing biological riddle and led to structural characterization of the rapid neurotransmitter release triggered by Syt1, a Ca^{2+} sensor that together with the SNARE complex (synaptobrevin, syntaxin-1 and SNAP-25) is responsible for this release [12]. The NMR data suggest a dynamic binding mode in which basic residues of the Syt1 C2B-domain β -sandwich bind to a polyacidic part of the SNARE complex (Fig. 22). Furthermore, Pan et al. demonstrated that it is possible to assign the methyl groups of the neuronal SNARE complex, i.e., the leucine, isoleucine, methionine and valine methyl groups, in a sequence-specific manner based solely on measurement of lanthanide-induced PCSs [169].

In 2015, Saio et al. reported that the detailed conformational characterization of the multidomain protein MurD with a molecular weight of 47 kDa was enabled by PCS analysis [170]. In order to detect sizeable PCSs, CLaNP-5 loaded with Yb^{3+} and Tm^{3+} was

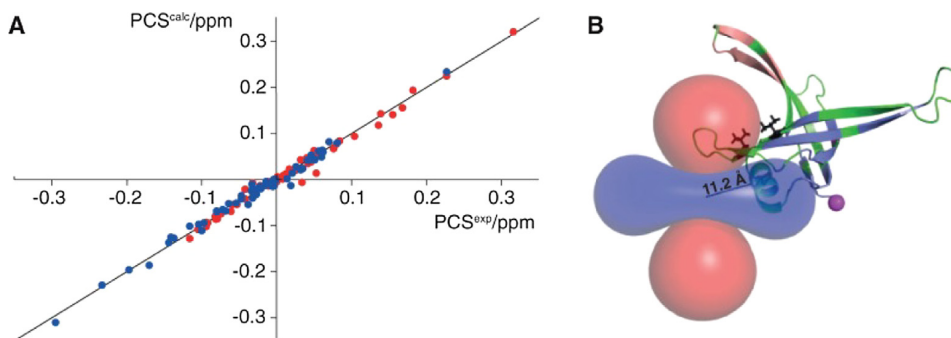


Fig. 21. PCS correlation plot measured on SSBP (A, red: Tb^{3+} , blue: Tm^{3+}). Orientation of the isosurfaces of the fitted anisotropy parameters shown together with the structure of SSBP (B) [165].

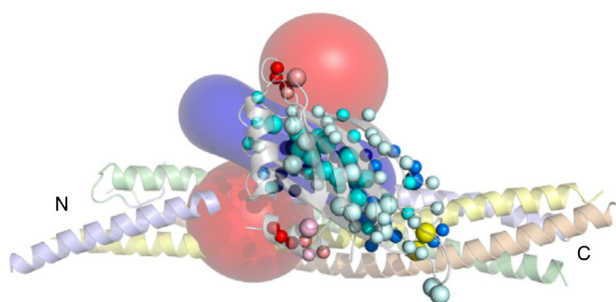


Fig. 22. Models of Syt C2B bound to the SNARE complex matching the Syt C2B PCSs induced by an LCT bound to residue 166 [12].

introduced into the protein by ligation of the cysteine residues of the mutant MurD E260C K262C. Upon analysis of the obtained structural restraints, the researchers found evidence for a novel semi-closed conformation that can exhibit different binding affinities compared to the open or closed conformation of the protein. The results are of high importance in the design of novel antibiotics by fragment based drug discovery, since, based on data from X-ray crystallography, MurD was previously thought to adopt only an open or closed conformation.

Pilla et al. demonstrated the use of LCTs for the generation of three-dimensional structure models of proteins larger than 200 residues relying solely on PCSs induced from 4 different tagging sites [103]. The new algorithm that they used relies on the Rosetta approach and yields fragment libraries that are then refined upon convergence. The generation of three-dimensional models of protein structures by PCSs and a Rosetta fragment library as the only input holds great promise for the structural investigation of proteins that are difficult to characterize by X-ray crystallography, cryo-EM and NOE-based methods.

Chen et al. found that for the bacteriophage T4 lysozyme (T4 Lys), which shows two subdomains that are arranged in two different conformations in its crystal structure, only one set of signals is present in NMR experiments, pointing to conformational homogeneity in solution [171]. In order to validate the results, a 4MDPA-derived LCT was attached to the protein. The obtained PCSs and RDCs strongly suggest that in absence of the substrate, the T4 lysozyme adapts, in contrast to the X-ray structure, a more open conformation.

Intrinsically disordered proteins (IDPs) are challenging subjects for structural characterization by NMR due to the significant overlap of resonances, as well as motions in the intermediate exchange regime, which can lead to signal broadening. In order to overcome severe overlap of signals, in 2016 Göbl et al. successfully demonstrated the use of an Yb³⁺ loaded DOTA-M8 LCT and were thereby able to increase the chemical shift dispersion, greatly facilitating analysis of the otherwise overlapped spectrum [73] (Fig. 23).

Although low-populated excited states can yield valuable information about a protein's function, it can be a challenging task to monitor these low abundance states and derive structural models from the data set. Ma et al. demonstrated that PCSs acquired via chemical exchange saturation transfer (CEST) provide a useful tool for both the acquisition of the signals of the otherwise invisible state as well as the structural modelling of the protein conformation [172]. The authors applied their approach during the characterization of the bound state of 3 mol% Ark1p peptides to the Abp1p SH3domain and the detection of transitional conformations of the FF domain.

In order to investigate the potential of LCTs for the determination of three-dimensional structures of proteins, Pearce et al. demonstrated that with only two different tagging sites and a sin-

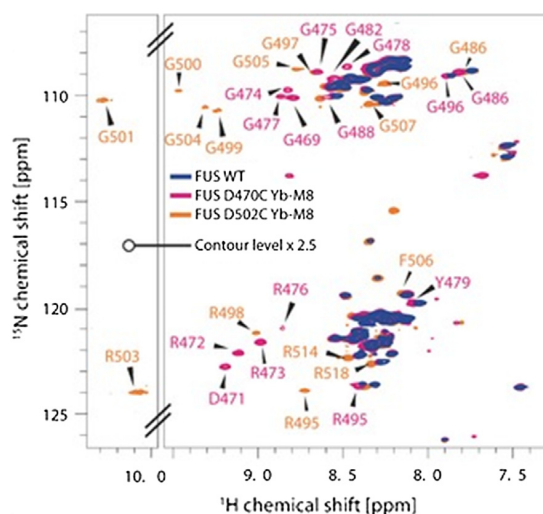


Fig. 23. Increasing the chemical shift dispersion of IDPs using an Yb-DOTA-M8 LCT as demonstrated by Göbl et al. (spectra acquired at 900 MHz proton frequency) [73].

gle lanthanide ion, namely Tb³⁺, the three-dimensional structure of the model protein ubiquitin can be investigated [26]. The authors of the study found that a previously published ensemble structure of ubiquitin fits better to the paramagnetic restraints than structures with only one conformation, a result that points out that PCSs can be used to detect very subtle changes in the conformation of biomacromolecules.

Lescanne et al. reported the assignment of methyl groups within the 25 kDa N-terminal domain of Hsp90 achieved by the analysis of PCSs induced by Yb³⁺-loaded ClANP-5 [24] from two or three different tagging sites using a novel version of the program PARAssign [173]. 60% of the 76 methyl groups were assigned in reliable manner, while for the majority of the remaining 40% only three or less potential assignments remained. In order to provide a complete assignment of the methyl groups, the data obtained from PCS analysis can be combined with additional restraints, e.g., from NOE experiments or mutagenesis studies. In a further study, the authors demonstrated that the PCS of the methyl groups can then be used to monitor the reorientation of methyl groups upon binding of a 4-(2-fluorophenyl)-2-pyrimidinamine ligand [174]. By application of their method, the authors found displacements of 1–3 Å for the methyl groups in the binding site and thereby showed that conformational changes of the side chains within a protein can be monitored using solely two-dimensional ¹H-¹³C HSQC spectra.

3.2. Applications of lanthanide chelating tags to proteins II: residual dipolar couplings

In 2007, Kamen et al. reported the use of an LCT for the characterization of membrane proteins [175]. Since the NMR signals are significantly broadened due to the large size of the protein and the use of a phospholipid bicelle or a solubilizing detergent micelle, deuteration of the investigated protein is required so that NOE-derived structures rely only on NOE signals of back-exchanged backbone amide protons. While use of polyacrylamide gels reduces the number of observed peaks and introduces additional line-broadening, EDTA-derived LCTs offer a robust method for partial alignment of the investigated membrane proteins and a successful measurement of ¹H-¹⁵N, ¹³C-¹⁵N and ¹³C-¹³C^α RDCs in 3D-experiments was demonstrated.

Xu et al. presented the use of an LCT attached to yeast cytochrome *c* in the study of a cytochrome *c*-adrenodoxin complex [176]. The CLaNP-5 tag did not interfere with the binding and showed the highly dynamical nature of the formed protein complex. While the detected RDCs for cytochrome *c* led to successful determination of a magnetic susceptibility tensor, the observed RDCs on adrenodoxin were significantly smaller than expected, indicating multiple conformations and dynamical averaging.

Although experimental evidence from cryo-EM and X-ray crystallography suggests a homotrimeric, C_3 symmetric structure of the transmembrane domain of the HIV-1 viral envelope protein gp41, Chiliveri et al. showed by the use of RDCs observed under three different alignment conditions (neutral and positively charged acrylamide gels as well as paramagnetic alignment using a thulium-containing LCT) and DEER EPR, that in a bicelle system the transmembrane domain is monomeric and exhibits a highly ordered element of 32 residues [177] (Fig. 24).

In 2019, Barnes et al. showed by introduction of a thulium-loaded DOTA-derived LCT that the 68-residue long medial-tail domain of myosin-VI, which corresponds to a highly charged single α -helical domain (SAH domain), adopts the structure of a highly ordered α -helix [178] in solution. Residues more distant from the attachment site show a decrease in alignment and these values were used for an exact calculation of the persistence length (defined as the distance in a polymer upon which correlations are lost) of the SAH helix. The result of 224 Å confirms the hypothesis that the investigated SAH helix is sufficiently stiff to be a part of the lever arm of myosin-VI.

3.3. Applications of lanthanide chelating tags to proteins III: protein-ligand complexes

Although the localization of ligands within proteins is highly important for any studies relating to drug molecules and their associated mechanism, it is still not straightforward today to localize ligands on proteins in complexes that are not suitable for crystallization and thus not readily accessible to X-ray crystallography. Ligand positions can be identified using NMR techniques such as CSP or NOESY experiments. While CSP data can be misleading or yield ambiguous results due to conformational changes occurring

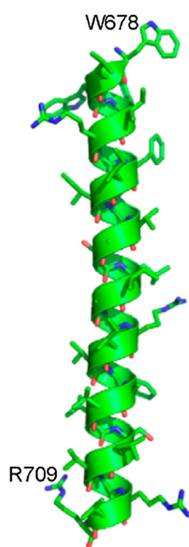


Fig. 24. Model of the transmembrane domain of the HIV-1 viral envelope protein gp41 modelled using RDCs by Chiliveri et al. [177]. (Reproduced with permission from Ref. [177]. Copyright 2018 American Chemical Society.)

in the protein upon ligand binding or unexpected intermolecular interactions, NOESY spectra and their analysis require in most cases a complete NMR assignment of the suitably labelled protein of interest. PCS analysis can provide the needed 3D information in a GPS-like approach from simple one-dimensional NMR experiments.

In 2011, Saio et al. showed that localization of ligands on proteins labelled with an LCT is possible by using structural restraints obtained by PCS [70]. Upon labelling of the protein of interest and determination of the magnetic susceptibility tensors associated with different lanthanide ions, the observed PCSs on the ligand atoms were translated into distance restraints. In combination with a docking calculation, Saio et al. successfully localized a high-affinity inhibitor within Grb2 SH2 over a distance of ~ 10 Å.

In contrast to strongly binding ligands, weakly binding ligands can be added in excess and hence the signal discrimination between protein and ligand is more conveniently achieved than in the case of strong binders. From the fraction of bound ligand, the PCSs can then be extracted. Guan et al. demonstrated this methodology for the localization of weakly binding ligands on a protein using two-point anchored LCTs [69] (Fig. 25). Due to their high degree of immobilization on the protein scaffold, two-armed lanthanide probes induce large PCSs and provide the additional advantage that the tensors of the anisotropic part of the magnetic susceptibility can be predicted computationally in order to yield a low-resolution model of the ligand localization, which can be used for example to discriminate between two binding sites within a protein. The study performed by Guan et al. relied only on one-dimensional ^1H spectra and the observed PCSs were used as docking restraints.

In 2015, Brath et al. presented an approach for the localization of ligands within proteins that relies on attachment of the LCT to the ligand [67]. The applied methodology presents several advantages: (i) the solubility in an aqueous solution of hydrophobic ligands can be enhanced to facilitate the study of their interactions with proteins, (ii) engineering of protein cysteine mutants is not needed, (iii) multiple binding sites can easily be detected, (iv) if the ligand is sufficiently large, it can be ensured that no structural changes of the protein or the protein-ligand binding complex occur, so that the interactions of the ligand with the native protein can be detected. However, chemical modification/tagging of the ligand renders this method less suitable for screening of a large number of ligands within a protein, in contrast to the approach of tagging of the protein that is able to screen a large number of ligands with one tagged protein construct.

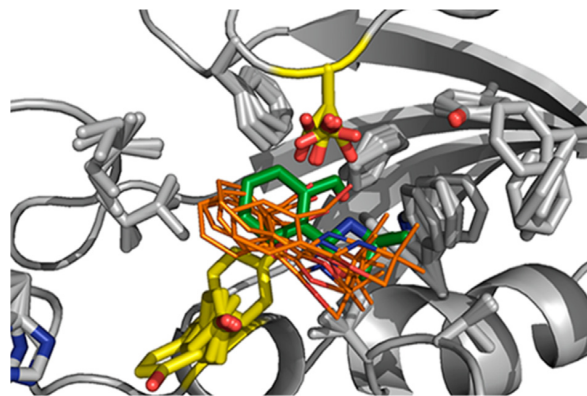


Fig. 25. Localization of a ligand on a protein by PCS reported by Guan et al. (green: averaged NOE structure, orange: best five structures obtained by PCS-fitting) [69]. (Reproduced with permission from Ref. [69]. Copyright 2013 American Chemical Society.)

Chen et al. reported in 2016 an interesting approach to overcome the problem of discrimination between signals of a strongly binding ligand and the protein [179]. A *tert*-butyl group was used as a reporter group attached to the ligand, since the *tert*-butyl group shows up in spectra as a relatively intense and extremely narrow signal with a specific chemical shift in the region of 1.0–1.5 ppm. Furthermore, the authors of the study found that the *tert*-butyl group gives rise to significant signals in NOESY spectra, even when protein NOESY correlations are hardly detectable. This approach is very convenient for the investigation of complexes of strongly binding, *tert*-butyl group containing ligands and a protein of interest.

Gao et al. demonstrated in 2017 the use of a ^{19}F -PCS CEST approach in order to characterize protein-ligand interactions in the intermediate exchange regime [66]. By scanning through ^{19}F frequencies while measuring the free ligand signal, intensity dips can be measured at the frequencies of the ^{19}F resonances in the bound state. Using this method, the researchers localized mono- and di-fluorinated inhibitors of the BRM bromodomain by PCS combined with docking approaches and confirmed the results by analysis of the solved crystal structure. Importantly, following initial determination of the magnetic susceptibility tensors, a large number of ligands can be screened without further need for protein engineering.

In 2018, Xu et al. presented a method suitable for the localization of ligands exhibiting micromolar affinities [180]. Since ligands with such affinities can fall in the range of intermediate exchange, and since ^1H CEST usually suffers from NOE interferences, the authors used a relaxation dispersion approach. The researchers successfully observed the desired PCSs and used them as docking restraints for the generation of structures of the protein-ligand complexes.

Zimmermann et al. reported that ^{19}F PCS NMR can be used as a method to determine fluorine positions of high affinity ligands located within hCA II accurately to within 0.8 Å [71] (Fig. 26). Upon determination of 4–5 independent magnetic susceptibility tensors induced by LCTs attached to different cysteine mutants, least-squares minimization and a Monte-Carlo approach allowed one to localize the ligands within the protein. The unprecedented distance range of 22–38 Å found for these effects holds great promise for investigation of protein-ligand complexes significantly larger than 30 kDa.

3.4. Applications of lanthanide chelating tags to carbohydrates

Since in sugar molecules, severe signal overlaps of virtually identical domains are frequently encountered, and often renders conventional J-coupling and NOE-based NMR techniques difficult to apply, introduction of LCTs and subsequent analysis of PCSs and RDCs can increase the shift dispersion and yield structural restraints that result in three-dimensional models and complement modelling studies.

In 2011, the first rigid LCT that was covalently attached to a carbohydrate molecule was reported by Mallagaray et al. [181]. In their study, the researchers demonstrated the successful transfer of the methodology, which was originally developed for proteins, by covalent attachment of an EDTA-based LCT via a rigid linker to 1- β -aminochitobiose and subsequent analysis of the observed PCSs.

Erdélyi et al. extended the methodology by attaching an EDTA-based chelator loaded with six different lanthanides to a lactose molecule, enabling an analysis of the conformational space of lactose [182]. The obtained restraints from RDCs and PCSs agree with molecular dynamics (MD) simulations that predicted that lactose is present in several different conformations of the glycosidic bond

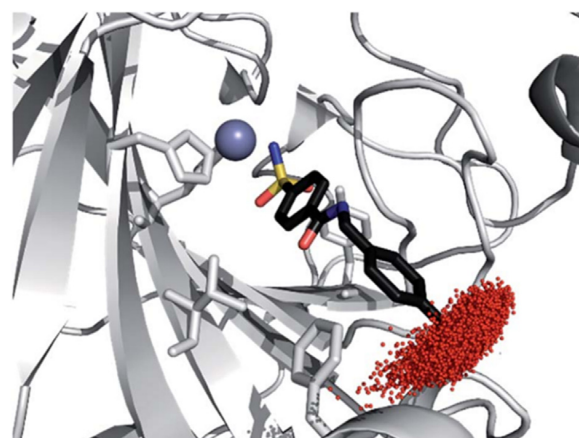


Fig. 26. Point cloud of the Monte-Carlo fluorine position calculation for FM-520 within hCA-II (light blue sticks (hidden by red point cloud): fluorine atoms, blue sphere: Zn^{2+} ion) [71]. (Reproduced with permission from Ref. [71].)

in an aqueous environment, and which exclude the possibility that lactose is present in solution in a single and rigid conformation.

The methodology was then extended by Yamamoto et al. to a typical core structure of N-linked oligosaccharides (N,N'-diacetyl chitobiose) [183]. Upon attachment of the EDTA-based LCT to the sugar molecule via a peptide coupling reaction employing HATU as coupling reagent, significant PCSs in ^1H - ^{13}C HSQC spectra were observed. The experimentally observed PCSs were used for a comparison with MD simulations and both approaches suggest that there is conformational rigidity of the glycosidic linkage in the investigated disaccharide and thus also in the common innermost part of N-linked oligosaccharides. The authors suggested that this method could be extended and interactions of N-linked oligosaccharides with lectins could be monitored using strongly paramagnetic lanthanides such as Dy^{3+} and Tb^{3+} .

The attachment of an LCT bound to a branched tetra-saccharide of biological importance, the branched tetra-saccharide of GM2 ganglioside, was presented by Zhang et al. [184] (Fig. 27). The combination of MD simulations and PCS analysis led to a model of the conformational space of GM2 tetra-saccharide. The comparison with its non-branched derivative, the GM3 trisaccharide, suggests that the additional GalNAc branch restricts the conformational freedom of the Neu5Ac-Gal glycosidic linkage in the GM2 tetra-saccharide.

Intricate N-glycans are important recognition molecules related to a diverse range of physiological pathways and diseases [185,186]. The structural investigation of the multiple and almost identical side-chains is challenging by standard methods such as non-paramagnetic NMR and X-ray crystallography. By introduction of an LCT, Canales et al. could distinguish the two arms of a symmetrical, branched N-glycan, which upon attachment of an LCT shows individual resonances for the two arms that exhibit otherwise isochronous signals, thereby confirming the major conformation and rotamer in solution [187]. Furthermore, this approach allowed characterization of the binding epitopes of the symmetrical glycan and showed that both arms of the N-glycan are involved in the interaction with human galectin-3. In a further study, Canales et al. showed that not only it is possible to determine carbohydrate conformations upon attachment of an LCT to a sugar molecule, but also to monitor the interaction site in a protein and characterize the topological changes upon binding of a model saccharide modified with an LCT [188]. Besides the fact that this method could also easily reveal secondary binding sites, the chemical modification of the sugar molecule with an LCT removes

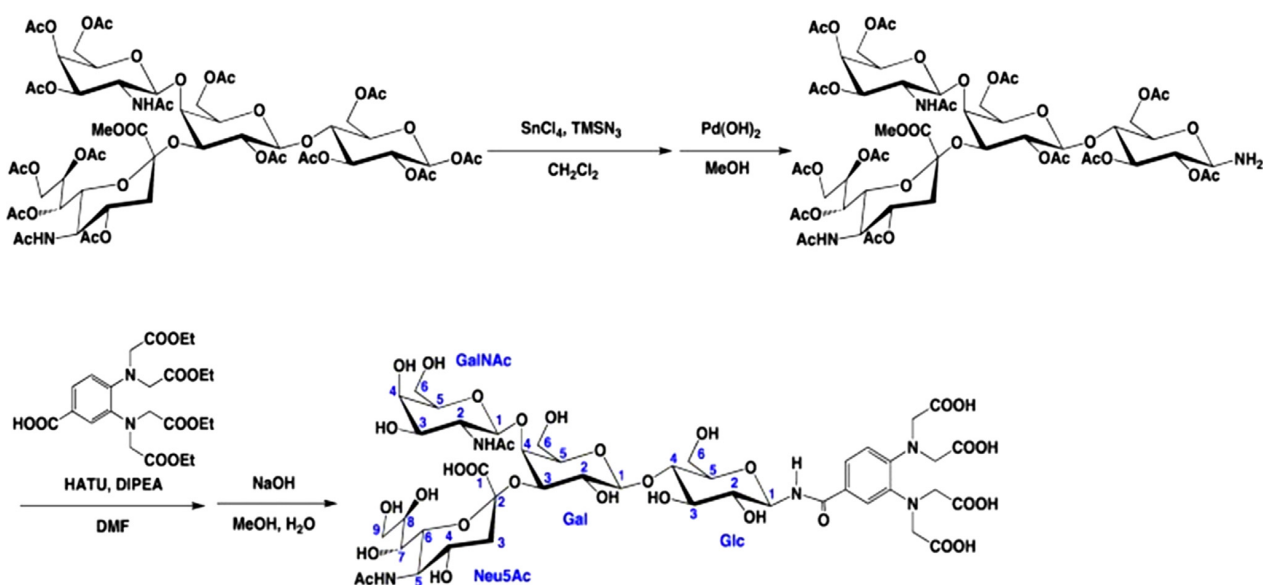


Fig. 27. Preparation of the branched tetra-saccharide covalently bound to a chelator reported in Zhang et al. [184].

the need for protein engineering. Canales et al. demonstrated that paramagnetic NMR allows one to increase the chemical shift dispersion of the virtually identical side-chains and the identification of the four N-acetyl lactosamine antennae recognized by two N-acetyllactosamine-binding lectins, specifically *Datura stramonium* seed lectin and *Ricinus communis* agglutinin [74]. After successful separation of the isochronous signals by PCS (Fig. 28) followed by structural characterization, addition of the lectin allowed the assignment of the major lectin-carbohydrate interactions by observation of the line broadening on the different branches of the tetra-antennary N-glycans.

Mallagaray et al. ligated an LCT via a copper-catalysed azide-alkyne click reaction to *L*-fucose [189]. Interestingly, the paramagnetic carbohydrate-LCT conjugate can then be used to aid the backbone assignment of the protruding domains of a human norovirus strain that constitutes a challenging target for investigation by NMR spectroscopy due to its high molecular weight of 70 kDa. The researchers found that the carbohydrate-LCT conjugate at low concentrations exclusively binds to the primary binding pocket. However, since peak broadening of multiple resonances was observed and compromised its use for the backbone assignment of this challenging target, synthesis of ligands with higher

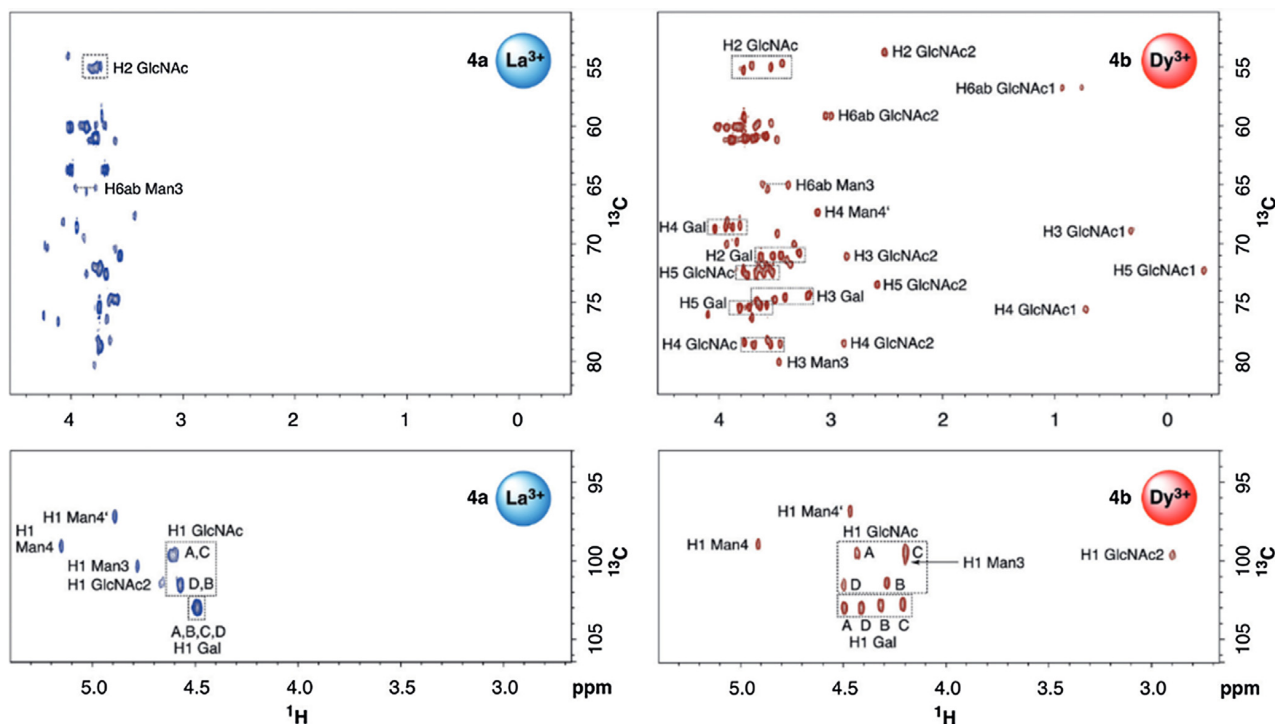


Fig. 28. Increasing the chemical shift dispersion in ^1H - ^{13}C HSQC spectra for the isochronous signals of the four N-acetyl lactosamine antennae as demonstrated by Canales et al. (acquired at 600 MHz proton frequency) [74].

binding affinity or the use of covalently bound tags on the biomacromolecule scaffold are highly desirable.

3.5. Applications of lanthanide chelating tags to nucleic acids

The structure of nucleic acids in solution is essential to their function and stability [190]. In order to investigate their structural properties, NMR offers a non-destructive approach to measure the solution state structure of DNA and RNA that not only allows characterization of these biomacromolecules in physiological conditions, but also tuning of the buffer salt concentration, pH value and additives in virtually any combination [191,192]. The wide range of LCTs that was developed for the investigation of structure and dynamics of proteins includes a large variety of different attachment systems (linker moieties forming disulphide bonds [9,24], bio-orthogonal linkers for ligation to unnatural amino acids [6], as well as reductively stable linkers for in-cell NMR [2–4]). Extension of this toolbox by new linkage systems or exploitation of the currently developed LCTs in order to study the structural properties of DNA and RNA is an important goal for future developments in the research field of paramagnetic NMR. By paramagnetic labelling of nucleic acids and derivatives, not only is it possible to obtain structural restraints for the assessment of computational studies and carry out refinement of structures obtained by X-ray crystallography and *de novo*-derived structures, it is also possible to derive ligand binding positions and characterize ligand-nucleic acid complexes.

In 1997, Gochin et al. presented the first application of paramagnetic NMR to nucleic acids, when they investigated the formation of a ternary drug-DNA-divalent cation complex composed of a nucleic acid strand, two chromomycin-A3 molecules and a (paramagnetic) Co^{2+} or a (diamagnetic) Zn^{2+} ion [193]. Subsequently, the signals were assigned and the magnetic susceptibility tensor parameters of the Co^{2+} complex were derived [194]. Furthermore, the structural restraints gained by analysis of the PCSs were used in energy minimization and MD calculations and shown to outperform structural restraints obtained from NOESY [29].

Further initial examples of the application of paramagnetic NMR applied to nucleic acids and their derivatives include the Cys-Ph-TAHA LCT, that was covalently attached to a ternary complex of the protein LacR, the O2 operator and IPTG in 2011 by Peters et al. [14] and the structural investigation of $\Sigma\text{ubp2-R3H}$ in complex with deoxyguanosine 5'-monophosphate (dGMP) as structural analogue to the 5' end of single-stranded DNA in 2012 by Jaudzems et al. [195]. In the latter example, paramagnetic

NMR provided insight into structural changes of the protein upon binding of the DNA mimic. In 2019 Strickland et al. reported that attachment of DOTA-M8-SSPy LCTs to a spliceosomal U1A RNA-binding domain allows structural characterization of a 232-nt RNA derived from the HIV-1 Rev response element [196]. Furthermore, the method was validated using a 36-nt RNA with known structural properties. Based on the fact that the seven nucleotide long U1A binding sites can be readily inserted in hairpin or loop structures, the authors suggest that the presented method is generally applicable for the atomic-level study of large RNAs.

These examples differ in the mode of the LCT-biomacromolecule linkage; in the first three examples affinity binding of the paramagnetic metal centre is exploited, whereas Peters et al., Jaudzems et al. and Strickland et al. attached the LCT covalently to the protein interacting with a nucleic acid or a nucleic acid derivative.

In order to minimize motional averaging of the paramagnetic effects observed in affinity binding, the covalent attachment of the LCT incorporating the paramagnetic metal centre directly to the nucleic acids is highly desirable. The covalent labelling of nucleic acids was demonstrated for an EPR application by Song et al. [197] and Goldfarb [198] via a copper catalysed azide-alkyne click reaction of a DNA strand incorporating an azide functionality and an LCT bearing an alkyne functionality, as well as by Wu et al. [30] for the analysis of PCSs via modification of a phosphorothioate group in DNA with an LCT containing an electrophilic bromoacetamide functionality for alkylation (Fig. 29).

Insights into the solution structure and dynamics of RNA and DNA molecules obtained by paramagnetic NMR are highly desirable not only for fundamental research, but also because they can give significant inputs for characterization of binding sites of molecules that target RNA and DNA important in future drug discovery. The information obtained by paramagnetic NMR can refine X-ray structures as well as complement or support modelling studies of the structure of nucleic acids. Since the application of LCTs for proteins is well established and documented, it seems highly promising to transfer this knowledge to applications involving nucleic acids in the next decade. Crucial factors for a successful transfer of these tools from proteins to nucleic acids will be the development of synthetically accessible LCTs bearing linkers suitable for a covalent attachment to nucleic acids. The structural insights gained by the exploitation of the paramagnetic effects from LCTs can then provide important data for fundamental research as well as catalyse the development of DNA and RNA targeting treatments in drug discovery.

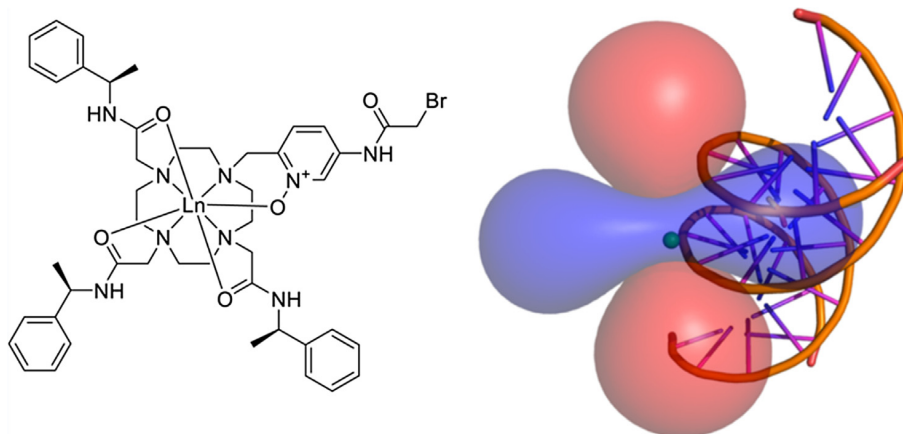


Fig. 29. Structure of the bromoacetamide LCT (left) employed by Wu et al. and isosurfaces of the anisotropy parameters obtained after successful ligation of the LCT to a phosphorothioate modified nucleophilic site in DNA (right) [30]. (Reproduced with permission from Ref. [30]. Copyright 2017 American Chemical Society.)

4. Conclusion & outlook

To conclude, over the last two decades a huge variety of LCTs has been developed, each with its own characteristics and fields of application. In contrast to the initial developments in the field that included LCTs whose use was hampered by multiple sets of signals and strongly averaged paramagnetic effects, the latest generations of LCTs available today use reduction-stable linkers and display only a single set of peaks. Recently published tags show very large anisotropy parameters ($\Delta\chi_{\text{ax}} > 40 \times 10^{-32} \text{ m}^3$) and consequently provide extremely large PCSs and RDCs, giving rise to precise structural restraints and convenient analysis of the acquired spectra. If one assumes a detection limit of 0.005 ppm for chemical shift differences, current LCTs potentially provide structural restraints over distances exceeding 200 Å. In combination with a sufficient number of suitable protein constructs to achieve the determination of the position of a spin with an accuracy within 0.8 Å, this implies that PCSs have become a very powerful tool for future applications. The possible spectrum of applications of LCTs is extremely broad and includes the investigation of protein structures in solution, protein structures in living cells, the structural characterization of flexible domains and conformations, the identification and investigation of ligand binding sites, the structure elucidation of protein-protein complexes, as well as applications on other biomacromolecules, such as DNA, RNA and oligosaccharides. Most importantly, LCTs can aid the refinement of structures of biomacromolecules in solution under physiological conditions and, as demonstrated in various studies, can lead to completely new and surprising findings that could not be revealed by X-ray structures or NOE studies. By giving a more detailed mechanistic insight into enzyme function and the alignment of protein subunits in solution, lanthanide-chelating tags have significantly contributed to the current knowledge of the structures and interactions of biomacromolecules and can lead to breakthroughs in drug development. In the near future, further developments of LCTs with suitable linkers for the investigation of DNA and RNA are highly desirable, as well as applications of existing LCTs to solve long-standing problems in structural biology, especially those associated with mechanisms of disease and biomedical applications. Open questions about the design of LCTs, however, remain. Thus, there is no systematic study so far about the influence of the set of donor atoms on the anisotropy parameters; the motional degrees of freedom for a given tag are still not entirely understood, and many alternative chelators to DOTA as basic scaffold have not yet been thoroughly tested for PCS – so there is still a great deal of work to be done.

Declaration of competing interest

The authors declared that there is no conflict of interest.

Acknowledgements

The Chemistry Department of the University of Basel is acknowledged for financial support. T. Müntener and R. Vogel are acknowledged for helpful discussions.

References

- [1] C. Nitsche, G. Otting, Pseudocontact shifts in biomolecular NMR using paramagnetic metal tags, *Prog. Nucl. Magn. Reson. Spectrosc.* 98–99 (2017) 20–49.
- [2] B.-B. Pan, F. Yang, Y. Ye, Q. Wu, C. Li, T. Huber, X.-C. Su, 3D structure determination of a protein in living cells using paramagnetic NMR spectroscopy, *Chem. Commun.* 52 (2016) 10237–10240.
- [3] T. Müntener, D. Häussinger, P. Selenko, F.-X. Theillet, In-Cell Protein Structures from 2D NMR Experiments, *J. Phys. Chem. Lett.* 7 (2016) 2821–2825.
- [4] Y. Hikone, G. Hirai, M. Mishima, K. Inomata, T. Ikeya, S. Arai, M. Shirakawa, M. Sodeoka, Y. Ito, A new carbamidemethyl-linked lanthanoid chelating tag for PCS NMR spectroscopy of proteins in living HeLa cells, *J. Biomol. NMR* 66 (2016) 99–110.
- [5] W.-M. Liu, M. Overhand, M. Ubbink, The application of paramagnetic lanthanoid ions in NMR spectroscopy on proteins, *Coord. Chem. Rev.* 273–274 (2014) 2–12.
- [6] C.T. Loh, K. Ozawa, K.L. Tuck, N. Barlow, T. Huber, G. Otting, B. Graham, Lanthanide Tags for Site-Specific Ligation to an Unnatural Amino Acid and Generation of Pseudocontact Shifts in Proteins, *Bioconjugate Chem.* 24 (2013) 260–268.
- [7] G. Otting, Protein NMR using Paramagnetic Ions, *Annu. Rev. Biophys.* 39 (2010) 387–405.
- [8] X.-C. Su, H. Liang, K.V. Loscha, G. Otting, [Ln(DPA)3]3– Is a Convenient Paramagnetic Shift Reagent for Protein NMR Studies, *J. Am. Chem. Soc.* 131 (2009) 10352–10353.
- [9] D. Häussinger, J.-R. Huang, S. Grzesiek, DOTA-M8: an extremely rigid, high-affinity lanthanide chelating tag for PCS NMR spectroscopy, *J. Am. Chem. Soc.* 131 (2009) 14761–14767.
- [10] X.-C. Su, B. Man, S. Beeren, H. Liang, S. Simonsen, C. Schmitz, T. Huber, B.A. Messerle, G. Otting, A dipicolinic acid tag for rigid lanthanide tagging of proteins and paramagnetic NMR spectroscopy, *J. Am. Chem. Soc.* 130 (2008) 10486–10487.
- [11] G. Pintacuda, A.Y. Park, M.A. Keniry, N.E. Dixon, G. Otting, Lanthanide labeling offers fast NMR approach to 3D structure determinations of protein–protein complexes, *J. Am. Chem. Soc.* 128 (2006) 3696–3702.
- [12] K.D. Brewer, T. Bacaj, A. Cavalli, C. Camilloni, J.D. Swarbrick, J. Liu, A. Zhou, P. Zhou, N. Barlow, J. Xu, A.B. Seven, E.A. Prinslow, R. Voletti, D. Häussinger, A.M.J. J. Bonvin, D.R. Tomchick, M. Vendruscolo, B. Graham, T.C. Südhof, J. Rizo, Dynamic binding mode of a Synaptotagmin-1–SNARE complex in solution, *Nat. Struct. Mol. Biol.* 22 (2015) 555.
- [13] V. Gaponenko, A.S. Altieri, J. Li, R.A. Byrd, Breaking symmetry in the structure determination of (large) symmetric protein dimers, *J. Biomol. NMR* 24 (2002) 143–148.
- [14] F. Peters, M. Maestre-Martinez, A. Leonov, L. Kovacic, S. Becker, R. Boelens, C. Griesinger, Cys-Ph-TAHA: a lanthanide binding tag for RDC and PCS enhanced protein NMR, *J. Biomol. NMR* 51 (2011) 329–337.
- [15] J. Wöhrnt, K.J. Franz, M. Nitz, B. Imperiali, H. Schwalbe, Protein alignment by a coexpressed lanthanide-binding tag for the measurement of residual dipolar couplings, *J. Am. Chem. Soc.* 125 (2003) 13338–13339.
- [16] I. Bertini, P. Turano, A.J. Vila, Nuclear magnetic resonance of paramagnetic metalloproteins, *Chem. Rev.* 93 (1993) 2833–2932.
- [17] J. Iwahara, G.M. Clore, Detecting transient intermediates in macromolecular binding by paramagnetic NMR, *Nature* 440 (2006) 1227.
- [18] I. Bertini, C. Luchinat, G. Parigi, R. Pierattelli, NMR spectroscopy of paramagnetic metalloproteins, *ChemBioChem* 6 (2005) 1536–1549.
- [19] L. Banci, I. Bertini, K.L. Bren, M.A. Cremonini, H.B. Gray, C. Luchinat, P. Turano, The use of pseudocontact shifts to refine solution structures of paramagnetic metalloproteins: Met80Ala cyano-cytochrome c as an example, *J. Biol. Inorg. Chem.* 1 (1996) 117–126.
- [20] G. Pintacuda, M. John, X.-C. Su, G. Otting, NMR structure determination of protein–ligand complexes by lanthanide labeling, *Acc. Chem. Res.* 40 (2007) 206–212.
- [21] T. Ikegami, L. Verdier, P. Sakhaii, S. Grimme, B. Pescatore, K. Saxena, K.M. Fiebig, C. Griesinger, Novel techniques for weak alignment of proteins in solution using chemical tags coordinating lanthanide ions, *J. Biomol. NMR* 29 (2004) 339–349.
- [22] G.M. Clore, J. Iwahara, Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes, *Chem. Rev.* 109 (2009) 4108–4139.
- [23] D. Joss, R.M. Walliser, K. Zimmermann, D. Häussinger, Conformationally locked lanthanide chelating tags for convenient pseudocontact shift protein nuclear magnetic resonance spectroscopy, *J. Biomol. NMR* (2018) 29–38.
- [24] P.H. Keizers, A. Saragliadis, Y. Hiruma, M. Overhand, M. Ubbink, Design, synthesis, and evaluation of a lanthanide chelating protein probe: ClaNp-5 yields predictable paramagnetic effects independent of environment, *J. Am. Chem. Soc.* 130 (2008) 14802–14812.
- [25] M.D. Lee, C.T. Loh, J. Shin, S. Chhabra, M.L. Dennis, G. Otting, J.D. Swarbrick, B. Graham, Compact, hydrophilic, lanthanide-binding tags for paramagnetic NMR spectroscopy, *Chem. Sci.* 6 (2015) 2614–2624.
- [26] B.J.G. Pearce, S. Jabar, C.-T. Loh, M. Szabo, B. Graham, G. Otting, Structure restraints from heteronuclear pseudocontact shifts generated by lanthanide tags at two different sites, *J. Biomol. NMR* 68 (2017) 19–32.
- [27] M. Regueiro-Figueroa, B. Bensenane, E. Ruscsák, D. Esteban-Gómez, L.J. Charbonnière, G. Tircsó, I. Tóth, A.D. Blas, T. Rodríguez-Blas, C. Platas-Iglesias, Lanthanide dota-like complexes containing a picolinate pendant: structural entry for the design of LnIII-based luminescent probes, *Inorg. Chem.* 50 (2011) 4125–4141.
- [28] M. Strickland, C.D. Schwieters, C. Göbl, A.C.L. Opina, M.-P. Strub, R.E. Swenson, O. Vasalatiy, N. Tjandra, Characterizing the magnetic susceptibility tensor of lanthanide-containing polymethylated-DOTA complexes, *J. Biomol. NMR* 66 (2016) 125–139.
- [29] K. Tu, M. Gochin, Structure determination by restrained molecular dynamics using NMR pseudocontact shifts as experimentally determined constraints, *J. Am. Chem. Soc.* 121 (1999) 9276–9285.

- [30] Z. Wu, M.D. Lee, T.J. Carruthers, M. Szabo, M.L. Dennis, J.D. Swarbrick, B. Graham, G. Otting, New Lanthanide tag for the generation of pseudocontact shifts in DNA by site-specific ligation to a phosphorothioate group, *Bioconjugate Chem.* 28 (2017) 1741–1748.
- [31] H. Yagi, K.V. Loscha, X.-C. Su, M. Stanton-Cook, T. Huber, G. Otting, Tunable paramagnetic relaxation enhancements by [Gd(DPA)3]³⁻ for protein structure analysis, *J. Biomol. NMR* 47 (2010) 143–153.
- [32] B. Liang, J.H. Bushweller, L.K. Tamm, Site-directed parallel spin-labeling and paramagnetic relaxation enhancement in structure determination of membrane proteins by solution NMR spectroscopy, *J. Am. Chem. Soc.* 128 (2006) 4389–4397.
- [33] J.R. Gillespie, D. Shortle, Characterization of long-range structure in the denatured state of staphylococcal nuclease. II. Distance restraints from paramagnetic relaxation and calculation of an ensemble of structures, *J. Mol. Biol.* 268 (1997) 170–184.
- [34] J. Iwahara, C.D. Schwieters, G.M. Clore, Ensemble approach for NMR structure refinement against ¹H paramagnetic relaxation enhancement data arising from a flexible paramagnetic group attached to a macromolecule, *J. Am. Chem. Soc.* 126 (2004) 5879–5896.
- [35] J. Iwahara, C. Tang, G. Marius Clore, Practical aspects of ¹H transverse paramagnetic relaxation enhancement measurements on macromolecules, *J. Magn. Reson.* 184 (2007) 185–195.
- [36] L.W. Donaldson, N.R. Skrynnikov, W.-Y. Choy, D.R. Muhandiram, B. Sarkar, J.D. Forman-Kay, L.E. Kay, Structural characterization of proteins with an attached ATCUN motif by paramagnetic relaxation enhancement NMR spectroscopy, *J. Am. Chem. Soc.* 123 (2001) 9843–9847.
- [37] G.M. Clore, M.R. Starich, A.M. Gronenborn, Measurement of residual dipolar couplings of macromolecules aligned in the nematic phase of a colloidal suspension of rod-shaped viruses, *J. Am. Chem. Soc.* 120 (1998) 10571–10572.
- [38] M.W.F. Fischer, J.A. Losonczy, J.L. Weaver, J.H. Prestegard, Domain orientation and dynamics in multidomain proteins from residual dipolar couplings, *Biochemistry* 38 (1999) 9013–9022.
- [39] J.R. Tolman, H.M. Al-Hashimi, L.E. Kay, J.H. Prestegard, Structural and dynamic analysis of residual dipolar coupling data for proteins, *J. Am. Chem. Soc.* 123 (2001) 1416–1424.
- [40] J. Meiler, J.J. Prompers, W. Peti, C. Griesinger, R. Brüschweiler, Model-free approach to the dynamic interpretation of residual dipolar couplings in globular proteins, *J. Am. Chem. Soc.* 123 (2001) 6098–6107.
- [41] M. Blackledge, Recent progress in the study of biomolecular structure and dynamics in solution from residual dipolar couplings, *Prog. Nucl. Magn. Reson. Spectrosc.* 46 (2005) 23–61.
- [42] P. Dossset, J.-C. Hus, D. Marion, M. Blackledge, A novel interactive tool for rigid-body modeling of multi-domain macromolecules using residual dipolar couplings, *J. Biomol. NMR* 20 (2001) 223–231.
- [43] W. Peti, J. Meiler, R. Brüschweiler, C. Griesinger, Model-free analysis of protein backbone motion from residual dipolar couplings, *J. Am. Chem. Soc.* 124 (2002) 5822–5833.
- [44] J.R. Tolman, K. Ruan, NMR residual dipolar couplings as probes of biomolecular dynamics, *Chem. Rev.* 106 (2006) 1720–1736.
- [45] P. Bernadó, C.W. Bertoncini, C. Griesinger, M. Zweckstetter, M. Blackledge, Defining long-range order and local disorder in native α -synuclein using residual dipolar couplings, *J. Am. Chem. Soc.* 127 (2005) 17968–17969.
- [46] K. Chen, N. Tjandra, The use of residual dipolar coupling in studying proteins by NMR, *Top. Curr. Chem.* 326 (2012) 47–67.
- [47] R. Barbieri, I. Bertini, G. Cavallaro, Y.-M. Lee, C. Luchinat, A. Rosato, Paramagnetically induced residual dipolar couplings for solution structure determination of lanthanide binding proteins, *J. Am. Chem. Soc.* 124 (2002) 5581–5587.
- [48] T. Saio, K. Ishimori, Accelerating structural life science by paramagnetic lanthanide probe methods, *Biochim. Biophys. Acta. Gen. Subj.* (2019), <https://doi.org/10.1016/j.bbagen.2019.03.018> (in press).
- [49] C.D. Barry, A.C.T. North, J.A. Glasel, R.J.P. Williams, A.V. Xavier, Quantitative determination of mononucleotide conformations in solution using lanthanide ion shift and broadening NMR probes, *Nature* 232 (1971) 236–245.
- [50] B. Bleaney, C.M. Dobson, B.A. Levine, R.B. Martin, R.J.P. Williams, A.V. Xavier, Origin of lanthanide nuclear magnetic resonance shifts and their uses, *J. Chem. Soc., Chem. Commun.* (1972) 791b–793b.
- [51] B. Bleaney, Nuclear magnetic resonance shifts in solution due to lanthanide ions, *J. Magn. Reson.* 8 (1972) 91–100.
- [52] A.M. Funk, K.-L.N.A. Finney, P. Harvey, A.M. Kenwright, E.R. Neil, N.J. Rogers, P. Kanthi Senanayake, D. Parker, Critical analysis of the limitations of Bleaney's theory of magnetic anisotropy in paramagnetic lanthanide coordination complexes, *Chem. Sci.* 6 (2015) 1655–1662.
- [53] E.A. Sutorina, K. Mason, C.F.G.C. Geraldes, I. Kuprov, D. Parker, Beyond Bleaney's theory: experimental and theoretical analysis of periodic trends in lanthanide-induced chemical shift, *Angew. Chem. Int. Ed.* 56 (2017) 12215–12218.
- [54] C.C. Hinckley, Paramagnetic shifts in solutions of cholesterol and the dipyrindine adduct of trisdipivalomethanatoeuropium (III). A shift reagent, *J. Am. Chem. Soc.* 91 (1969) 5160–5162.
- [55] P.V. Demarco, T.K. Elzey, R.B. Lewis, E. Wenkert, Tris(dipivalomethanato)europium (III). Shift reagent for use in the proton magnetic resonance analysis of steroids and terpenoids, *J. Am. Chem. Soc.* 92 (1970) 5737–5739.
- [56] A.F. Cockerill, D.M. Rackham, Quantitation of the chemical shifts induced by tris(dipivalomethanato)europium III in the PMR spectra of hydroxyadamantanes and cyclopentanol, *Tetrahedron Lett.* 11 (1970) 5149–5152.
- [57] G.M. Whitesides, D.W. Lewis, Tris[3-(tert-butylhydroxymethylene)-d-camphorato]europium(III). A reagent for determining enantiomeric purity, *J. Am. Chem. Soc.* 92 (1970) 6979–6980.
- [58] H.L. Goering, J.N. Eikenberry, G.S. Koerner, Tris[3-(trifluoromethylhydroxymethylene)-d-camphorato]europium(III). Chiral shift reagent for direct determination of enantiomeric compositions, *J. Am. Chem. Soc.* 93 (1971) 5913–5914.
- [59] T. Saio, K. Ogura, M. Yokochi, Y. Kobashigawa, F. Inagaki, Two-point anchoring of a lanthanide-binding peptide to a target protein enhances the paramagnetic anisotropic effect, *J. Biomol. NMR* 44 (2009) 157–166.
- [60] K.J. Franz, M. Nitz, B. Imperiali, Lanthanide-binding tags as versatile protein coexpression probes, *ChemBioChem* 4 (2003) 265–271.
- [61] M. Nitz, K.J. Franz, R.L. Maglathlin, B. Imperiali, A powerful combinatorial screen to identify high-affinity Terbium(III)-binding peptides, *ChemBioChem* 4 (2003) 272–276.
- [62] A. Leonov, B. Voigt, F. Rodriguez-Castañeda, P. Sakhaii, C. Griesinger, Convenient synthesis of multifunctional EDTA-based chiral metal chelates substituted with an S-mesylcysteine, *Chem. Eur. J.* 11 (2005) 3342–3348.
- [63] B. Graham, C.T. Loh, J.D. Swarbrick, P. Ung, J. Shin, H. Yagi, X. Jia, S. Chhabra, N. Barlow, G. Pintacuda, T. Huber, G. Otting, DOTA-amide lanthanide tag for reliable generation of pseudocontact shifts in protein NMR spectra, *Bioconjugate Chem.* 22 (2011) 2118–2125.
- [64] M. Gochin, H. Roder, Protein structure refinement based on paramagnetic NMR shifts: applications to wild-type and mutant forms of cytochrome c, *Protein Sci.* 4 (1995) 296–305.
- [65] Z. Gong, C.D. Schwieters, C. Tang, Theory and practice of using solvent paramagnetic relaxation enhancement to characterize protein conformational dynamics, *Methods* (2018).
- [66] J. Gao, E. Liang, R. Ma, F. Li, Y. Liu, J. Liu, L. Jiang, C. Li, H. Dai, J. Wu, X. Su, W. He, K. Ruan, Fluorine pseudocontact shifts used for characterizing the protein-ligand interaction mode in the limit of NMR intermediate exchange, *Angew. Chem. Int. Ed.* 56 (2017) 12982–12986.
- [67] U. Brath, S.I. Swamy, A.X. Veiga, C.-C. Tung, F. Van Petegem, M. Erdélyi, Paramagnetic ligand tagging to identify protein binding sites, *J. Am. Chem. Soc.* 137 (2015) 11391–11398.
- [68] M. John, G. Pintacuda, A.Y. Park, N.E. Dixon, G. Otting, Structure determination of protein–ligand complexes by transferred paramagnetic shifts, *J. Am. Chem. Soc.* 128 (2006) 12910–12916.
- [69] J.-Y. Guan, P.H.J. Keizers, W.-M. Liu, F. Löhr, S.P. Skinner, E.A. Heeneman, H. Schwalbe, M. Ubbink, G. Siegal, Small-molecule binding sites on proteins established by paramagnetic NMR spectroscopy, *J. Am. Chem. Soc.* 135 (2013) 5859–5868.
- [70] T. Saio, K. Ogura, K. Shimizu, M. Yokochi, T.R. Burke, F. Inagaki, An NMR strategy for fragment-based ligand screening utilizing a paramagnetic lanthanide probe, *J. Biomol. NMR* 51 (2011) 395.
- [71] K. Zimmermann, D. Joss, T. Müntener, E.S. Nogueira, M. Schäfer, L. Knörr, F.W. Monnard, D. Häussinger, Localization of ligands within human carbonic anhydrase II using ¹⁹F pseudocontact shift analysis, *Chem. Sci.* 10 (2019) 5064–5072.
- [72] C. Öster, S. Kosol, C. Hartlmüller, J.M. Lamley, D. Iuga, A. Oss, M.-L. Org, K. Vanatalu, A. Samoson, T. Madl, J.R. Lewandowski, Characterization of protein-protein interfaces in large complexes by solid-state NMR solvent paramagnetic relaxation enhancements, *J. Am. Chem. Soc.* 139 (2017) 12165–12174.
- [73] C. Göbl, M. Resch, M. Strickland, C. Hartlmüller, M. Viertler, N. Tjandra, T. Madl, Increasing the chemical-shift dispersion of unstructured proteins with a covalent lanthanide shift reagent, *Angew. Chem. Int. Ed.* 55 (2016) 14847–14851.
- [74] A. Canales, I. Boos, L. Perkams, L. Karst, T. Lubert, T. Karagiannis, G. Domínguez, F.J. Cañada, J. Pérez-Castells, D. Häussinger, C. Unverzagt, J. Jiménez-Barbero, Breaking the limits in analyzing carbohydrate recognition by NMR spectroscopy: resolving branch-selective interaction of a tetra-antennary N-glycan with lectins, *Angew. Chem. Int. Ed.* 56 (2017) 14987–14991.
- [75] M. Vonci, K. Mason, E.A. Sutorina, A.T. Frawley, S.G. Worswick, I. Kuprov, D. Parker, E.J.L. McInnes, N.F. Chilton, Rationalization of anomalous pseudocontact shifts and their solvent dependence in a series of C₃-symmetric lanthanide complexes, *J. Am. Chem. Soc.* 139 (2017) 14166–14172.
- [76] A. Roca-Sabio, M. Regueiro-Figueroa, D. Esteban-Gómez, A. de Blas, T. Rodríguez-Blas, C. Platas-Iglesias, Density functional dependence of molecular geometries in lanthanide(III) complexes relevant to bioanalytical and biomedical applications, *Comput. Theor. Chem.* 999 (2012) 93–104.
- [77] G. Hong, M. Dolg, L. Li, A comparison of scalar-relativistic ZORA and DKH density functional schemes: monohydrides, monooxides and monofluorides of La, Lu, Ac and Lr, *Chem. Phys. Lett.* 334 (2001) 396–402.
- [78] M. Purgel, Z. Baranyai, A. de Blas, T. Rodríguez-Blas, I. Bányai, C. Platas-Iglesias, I. Tóth, An NMR and DFT investigation on the conformational properties of Lanthanide(III) 1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetate analogues containing methylenephosphonate pendant arms, *Inorg. Chem.* 49 (2010) 4370–4382.
- [79] G. Schreckenbach, T. Ziegler, Calculation of the G-tensor of electron paramagnetic resonance spectroscopy using gauge-including atomic orbitals and density functional theory, *J. Phys. Chem. A* 101 (1997) 3388–3399.

- [80] S. Per, P.E.A. Turchi, A. Landa, V. Lordi, Ground-state properties of rare-earth metals: an evaluation of density-functional theory, *J. Phys. Condens. Matter* 26 (2014) 416001.
- [81] X. Dai, Y. Gao, M. Xin, Z. Wang, R. Zhou, The ground state and electronic structure of Gd@C82: a systematic theoretical investigation of first principle density functionals, *J. Chem. Phys.* 141 (2014) 244306.
- [82] S. Grimm, G. Schoendorff, A.K. Wilson, Gauging the performance of density functionals for lanthanide-containing molecules, *J. Chem. Theory Comput.* 12 (2016) 1259–1266.
- [83] L.E. Aebbersold, S.H. Yuwono, G. Schoendorff, A.K. Wilson, Efficacy of density functionals and relativistic effective core potentials for lanthanide-containing species: the Ln54 molecule set, *J. Chem. Theory Comput.* 13 (2017) 2831–2839.
- [84] L. Cerofolini, J.M. Silva, E. Ravera, M. Romanelli, C.F.G.C. Geraldes, A.L. Macedo, M. Fragai, G. Parigi, C. Luchinat, How do nuclei couple to the magnetic moment of a paramagnetic center? A new theory at the gauntlet of the experiments, *J. Phys. Chem. Lett.* 10 (2019) 3610–3614.
- [85] J. Koehler, J. Meiler, Expanding the utility of NMR restraints with paramagnetic compounds: Background and practical aspects, *Prog. Nucl. Magn. Reson. Spectrosc.* 59 (2011) 360–389.
- [86] F. Rodriguez-Castañeda, P. Haberz, A. Leonov, C. Griesinger, Paramagnetic tagging of diamagnetic proteins for solution NMR, *Magn. Reson. Chem.* 44 (2006) S10–S16.
- [87] X.-C. Su, G. Otting, Paramagnetic labelling of proteins and oligonucleotides for NMR, *J. Biomol. NMR* 46 (2010) 101–112.
- [88] P.H.J. Keizers, M. Ubbink, Paramagnetic tagging for protein structure and dynamics analysis, *Prog. Nucl. Magn. Reson. Spectrosc.* 58 (2011) 88–96.
- [89] K.N. Allen, B. Imperiali, Lanthanide-tagged proteins—an illuminating partnership, *Curr. Opin. Chem. Biol.* 14 (2010) 247–254.
- [90] X.-C. Su, J.-L. Chen, Site-specific tagging of proteins with paramagnetic ions for determination of protein structures in solution and in cells, *Acc. Chem. Res.* 52 (2019) 1675–1686.
- [91] M. John, G. Otting, Strategies for measurements of pseudocontact shifts in protein NMR spectroscopy, *ChemPhysChem* 8 (2007) 2309–2313.
- [92] G. Otting, Prospects for lanthanides in structural biology by NMR, *J. Biomol. NMR* 42 (2008) 1–9.
- [93] M.A.S. Hass, M. Ubbink, Structure determination of protein–protein complexes with long-range anisotropic paramagnetic NMR restraints, *Curr. Opin. Struct. Biol.* 24 (2014) 45–53.
- [94] C. Göbl, T. Madl, B. Simon, M. Sattler, NMR approaches for structural analysis of multidomain proteins and complexes in solution, *Prog. Nucl. Magn. Reson. Spectrosc.* 80 (2014) 26–63.
- [95] A. Carlon, E. Ravera, W. Andražojć, G. Parigi, G.N. Murshudov, C. Luchinat, How to tackle protein structural data from solution and solid state: an integrated approach, *Prog. Nucl. Magn. Reson. Spectrosc.* 92–93 (2016) 54–70.
- [96] A. Bhaumik, C. Luchinat, G. Parigi, E. Ravera, M. Rinaldelli, NMR crystallography on paramagnetic systems: solved and open issues, *CrytEngComm* 15 (2013) 8639–8656.
- [97] J.H. Prestegard, D.A. Agard, K.W. Moremen, L.A. Lavery, L.C. Morris, K. Pederson, Sparse labeling of proteins: structural characterization from long range constraints, *J. Magn. Reson.* 241 (2014) 32–40.
- [98] K. Kato, T. Yamaguchi, Paramagnetic NMR probes for characterization of the dynamic conformations and interactions of oligosaccharides, *Glycoconj. J.* 32 (2015) 505–513.
- [99] I. Bertini, C. Luchinat, G. Parigi, Magnetic susceptibility in paramagnetic NMR, *Prog. Nucl. Magn. Reson. Spectrosc.* 40 (2002) 249–273.
- [100] A.J. Pell, G. Pintacuda, C.P. Grey, Paramagnetic NMR in solution and the solid state, *Prog. Nucl. Magn. Reson. Spectrosc.* 111 (2019) 1–271.
- [101] C. Schmitz, M.J. Stanton-Cook, X.C. Su, G. Otting, T. Huber, Numbat: an interactive software tool for fitting Deltachi-tensors to molecular coordinates using pseudocontact shifts, *J. Biomol. NMR* 41 (2008) 179–189.
- [102] L. Yao, B. Vögeli, J. Ying, A. Bax, NMR determination of amide N–H equilibrium bond length from concerted dipolar coupling measurements, *J. Am. Chem. Soc.* 130 (2008) 16518–16520.
- [103] K.B. Pilla, G. Otting, T. Huber, Pseudocontact shift-driven iterative resampling for 3D structure determinations of large proteins, *J. Mol. Biol.* 428 (2016) 522–532.
- [104] G.M. Clore, Accurate and rapid docking of protein–protein complexes on the basis of intermolecular nuclear Overhauser enhancement data and dipolar couplings by rigid body minimization, *Proc. Natl. Acad. Sci. U.S.A.* 97 (2000) 9021.
- [105] J.H. Prestegard, H.M. Al-Hashimi, J.R. Tolman, NMR structures of biomolecules using field oriented media and residual dipolar couplings, *Q. Rev. Biophys.* 33 (2000) 371–424.
- [106] C.A. Bewley, G.M. Clore, Determination of the relative orientation of the two halves of the domain-swapped dimer of cyanovirin-N in solution using dipolar couplings and rigid body minimization, *J. Am. Chem. Soc.* 122 (2000) 6009–6016.
- [107] G.-W. Li, H. Liu, F. Qiu, X.-J. Wang, X.-X. Lei, Residual dipolar couplings in structure determination of natural products, *Nat. Prod. Bioprospect.* 8 (2018) 279–295.
- [108] D. MacDonald, P. Lu, Residual dipolar couplings in nucleic acid structure determination, *Curr. Opin. Struct. Biol.* 12 (2002) 337–343.
- [109] R.D. Shannon, Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides, *Acta Crystallogr. A* 32 (1976) 751–767.
- [110] I. Bertini, M.B.L. Janik, Y.-M. Lee, C. Luchinat, A. Rosato, Magnetic susceptibility tensor anisotropies for a lanthanide ion series in a fixed protein matrix, *J. Am. Chem. Soc.* 123 (2001) 4181–4188.
- [111] I. Bertini, A. Donaire, B. Jimenez, C. Luchinat, G. Parigi, M. Piccioli, L. Poggi, Paramagnetism-based versus classical constraints: an analysis of the solution structure of Ca Ln calbindin D9k, *J. Biomol. NMR* 21 (2001) 85–98.
- [112] C. Ma, S.J. Opella, Lanthanide ions bind specifically to an added “EF-Hand” and orient a membrane protein in micelles for solution NMR spectroscopy, *J. Magn. Reson.* 146 (2000) 381–384.
- [113] M. Nitz, M. Sherawat, K.J. Franz, E. Peisach, K.N. Allen, B. Imperiali, Structural origin of the high affinity of a chemically evolved lanthanide-binding peptide, *Angew. Chem. Int. Ed.* 43 (2004) 3682–3685.
- [114] T. Saio, M. Yokochi, H. Kumeta, F. Inagaki, PCS-based structure determination of protein–protein complexes, *J. Biomol. NMR* 46 (2010) 271–280.
- [115] Y. Kobashigawa, T. Saio, M. Ushio, M. Sekiguchi, M. Yokochi, K. Ogura, F. Inagaki, Convenient method for resolving degeneracies due to symmetry of the magnetic susceptibility tensor and its application to pseudo contact shift-based protein–protein complex structure determination, *J. Biomol. NMR* 53 (2012) 53–63.
- [116] X.-C. Su, K. McAndrew, T. Huber, G. Otting, Lanthanide-binding peptides for NMR measurements of residual dipolar couplings and paramagnetic effects from multiple angles, *J. Am. Chem. Soc.* 130 (2008) 1681–1687.
- [117] X.-C. Su, T. Huber, N.E. Dixon, G. Otting, Site-specific labelling of proteins with a rigid lanthanide-binding tag, *ChemBioChem* 7 (2006) 1599–1604.
- [118] M. Prudêncio, J. Rohovec, J.A. Peters, E. Tocheva, M.J. Boulanger, M.E.P. Murphy, H.-J. Hupkes, W. Koster, A. Impagliazzo, M. Ubbink, A caged lanthanide complex as a paramagnetic shift agent for protein NMR, *Chem. Eur. J.* 10 (2004) 3252–3260.
- [119] D. Parker, R.S. Dickens, H. Puschmann, C. Crossland, J.A.K. Howard, Being Excited by Lanthanide Coordination Complexes: Aqua Species, Chirality, Excited-State Chemistry, and Exchange Dynamics, *Chem. Rev.* 102 (2002) 1977–2010.
- [120] G. Pintacuda, A. Moshref, A. Leonchiks, A. Sharipo, G. Otting, Site-specific labelling with a metal chelator for protein-structure refinement, *J. Biomol. NMR* 29 (2004) 351–361.
- [121] A. Dvoretzky, V. Gaponenko, P.R. Rosevear, Derivation of structural restraints using a thiol-reactive chelator, *FEBS Lett.* 528 (2002) 189–192.
- [122] P. Haberz, F. Rodriguez-Castañeda, J. Junker, S. Becker, A. Leonov, C. Griesinger, Two new chiral EDTA-based metal chelates for weak alignment of proteins in solution, *Org. Lett.* 8 (2006) 1275–1278.
- [123] B. Man, X.-C. Su, H. Liang, S. Simonsen, T. Huber, B.A. Messerle, G. Otting, 3-Mercapto-2,6-pyridinedicarboxylic acid: a small lanthanide-binding tag for protein studies by NMR spectroscopy, *Chem. Eur. J.* 16 (2010) 3827–3832.
- [124] J.D. Swarbrick, P. Ung, X.-C. Su, A. Maleckis, S. Chhabra, T. Huber, G. Otting, B. Graham, Engineering of a bis-chelator motif into a protein α -helix for rigid lanthanide binding and paramagnetic NMR spectroscopy, *Chem. Commun.* 47 (2011) 7368–7370.
- [125] J.D. Swarbrick, P. Ung, S. Chhabra, B. Graham, An iminodiacetic acid based lanthanide binding tag for paramagnetic exchange NMR spectroscopy, *Angew. Chem. Int. Ed.* 50 (2011) 4403–4406.
- [126] H. Yagi, A. Maleckis, G. Otting, A systematic study of labelling an α -helix in a protein with a lanthanide using IDA-SH or NTA-SH tags, *J. Biomol. NMR* 55 (2013) 157–166.
- [127] X. Jia, A. Maleckis, T. Huber, G. Otting, 4,4'-Dithiobisdipicolinic acid: a small and convenient lanthanide binding tag for protein NMR spectroscopy, *Chem. Eur. J.* 17 (2011) 6830–6836.
- [128] Q.-F. Li, Y. Yang, A. Maleckis, G. Otting, X.-C. Su, Thiol–ene reaction: a versatile tool in site-specific labelling of proteins with chemically inert tags for paramagnetic NMR, *Chem. Commun.* 48 (2012) 2704–2706.
- [129] Y. Yang, Q.-F. Li, C. Cao, F. Huang, X.-C. Su, Site-specific labeling of proteins with a chemically stable, high-affinity tag for protein study, *Chem. Eur. J.* 19 (2013) 1097–1103.
- [130] F. Huang, Y.-Y. Pei, H.-H. Zuo, J.-L. Chen, Y. Yang, X.-C. Su, Bioconjugation of proteins with a paramagnetic NMR and fluorescent tag, *Chem. Eur. J.* 19 (2013) 17141–17149.
- [131] Y. Yang, J.-T. Wang, Y.-Y. Pei, X.-C. Su, Site-specific tagging proteins via a rigid, stable and short thioether tether for paramagnetic spectroscopic analysis, *Chem. Commun.* 51 (2015) 2824–2827.
- [132] C.-T. Loh, B. Graham, E.H. Abdalkader, K.L. Tuck, G. Otting, Generation of pseudocontact shifts in proteins with lanthanides using small “Clickable” nitrilotriacetic acid and iminodiacetic acid tags, *Chem. Eur. J.* 21 (2015) 5084–5092.
- [133] W.-X. Jiang, X.-H. Gu, X. Dong, C. Tang, Lanthanoid tagging via an unnatural amino acid for protein structure characterization, *J. Biomol. NMR* 67 (2017) 273–282.
- [134] J.-L. Chen, Y. Zhao, Y.-J. Gong, B.-B. Pan, X. Wang, X.-C. Su, Stable and rigid DTPA-like paramagnetic tags suitable for in vitro and in situ protein NMR analysis, *J. Biomol. NMR* 70 (2018) 77–92.
- [135] M.F. Tweedle, J.J. Hagan, K. Kumar, S. Mantha, C.A. Chang, Reaction of gadolinium chelates with endogenously available ions, *Magn. Reson. Imaging* 9 (1991) 409–415.
- [136] G.J. Stasiuk, N.J. Long, The ubiquitous DOTA and its derivatives: the impact of 1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid on biomedical imaging, *Chem. Commun.* 49 (2013) 2732–2746.

- [137] H.R. Maecke, J.C. Reubi, Somatostatin Receptors as Targets for Nuclear Medicine Imaging and Radionuclide Treatment, *J. Nucl. Med.* 52 (2011) 841–844.
- [138] D. Häussinger, S. Grzesiek, A new lanthanide-chelating tag with very high metal affinity was developed for protein NMR on a calcium binding protein, First European Conference on Chemistry for Life Sciences, Book of Abstracts, Rimini, Italy, 2005.
- [139] A.C.L. Opina, M. Strickland, Y.-S. Lee, N. Tjandra, R.A. Byrd, R.E. Swenson, O. Vasalati, Analysis of the isomer ratios of polymethylated-DOTA complexes and the implications on protein structural studies, *Dalton Trans.* 45 (2016) 4673–4687.
- [140] E.A. Sutura, D. Häussinger, K. Zimmermann, L. Garbuio, M. Yulikov, G. Jeschke, I. Kuprov, Model-free extraction of spin label position distributions from pseudocontact shift data, *Chem. Sci.* 8 (2017) 2751–2757.
- [141] V.S. Mironov, Y.G. Galyametdinov, A. Ceulemans, C. Görlner-Walrand, K. Binnemans, Influence of crystal-field perturbations on the room-temperature magnetic anisotropy of lanthanide complexes, *Chem. Phys. Lett.* 345 (2001) 132–140.
- [142] M.D. Lee, M.L. Dennis, B. Graham, J.D. Swarbrick, Short two-armed lanthanide-binding tags for paramagnetic NMR spectroscopy based on chiral 1,4,7,10-tetrakis(2-hydroxypropyl)-1,4,7,10-tetraazacyclododecane scaffolds, *Chem. Commun.* 53 (2017) 13205–13208.
- [143] M.D. Lee, M.L. Dennis, J.D. Swarbrick, B. Graham, Enantiomeric two-armed lanthanide-binding tags for complementary effects in paramagnetic NMR spectroscopy, *Chem. Commun.* 52 (2016) 7954–7957.
- [144] T. Müntener, J. Kottelat, A. Huber, D. Häussinger, New lanthanide chelating tags for PCS NMR spectroscopy with reduction stable, rigid linkers for fast and irreversible conjugation to proteins, *Bioconjugate Chem.* 29 (2018) 3344–3351.
- [145] D. Joss, M.-S. Bertrams, D. Häussinger, A sterically overcrowded, isopropyl-substituted lanthanide chelating tag for protein PCS NMR spectroscopy: Synthesis of its macrocyclic scaffold and benchmarking on ubiquitin S57C and hCA II S166C, *Chem. Eur. J.* (2019), <https://doi.org/10.1002/chem.201901692> (in press).
- [146] E.A. Sutura, I. Kuprov, Pseudocontact shifts from mobile spin labels, *Phys. Chem. Chem. Phys.* 18 (2016) 26412–26422.
- [147] D. Joss, D. Häussinger, P4T – A lanthanide chelating tag combining a highly sterically overcrowded backbone with a reductively stable linker, *Chem. Commun.* 55 (2019) 10543–10546, <https://doi.org/10.1039/c9cc04676c> (in press).
- [148] F. Yang, X. Wang, B.-B. Pan, X.-C. Su, Single-armed phenylsulfonated pyridine derivative of DOTA is rigid and stable paramagnetic tag in protein analysis, *Chem. Commun.* 52 (2016) 11535–11538.
- [149] P.H.J. Keizers, J.F. Desreux, M. Overhand, M. Ubbink, Increased paramagnetic effect of a lanthanide protein probe by two-point attachment, *J. Am. Chem. Soc.* 129 (2007) 9292–9293.
- [150] M.D. Vlasie, C. Comuzzi, A.M.C.H. van den Nieuwendijk, M. Prudêncio, M. Overhand, M. Ubbink, Long-range-distance NMR effects in a protein labeled with a lanthanide–DOTA chelate, *Chem. Eur. J.* 13 (2007) 1715–1723.
- [151] W.-M. Liu, S.P. Skinner, M. Timmer, A. Blok, M.A.S. Hass, D.V. Filipov, M. Overhand, M. Ubbink, A two-armed lanthanoid-chelating paramagnetic NMR probe linked to proteins via thioether linkages, *Chem. Eur. J.* 20 (2014) 6256–6258.
- [152] W.-M. Liu, P.H.J. Keizers, M.A.S. Hass, A. Blok, M. Timmer, A.J.C. Sarris, M. Overhand, M. Ubbink, A pH-sensitive, colorful, lanthanide-chelating paramagnetic NMR probe, *J. Am. Chem. Soc.* 134 (2012) 17306–17313.
- [153] D. Shishmarev, G. Otting, How reliable are pseudocontact shifts induced in proteins and ligands by mobile paramagnetic metal tags? A modelling study, *J. Biomol. NMR* 56 (2013) 203–216.
- [154] M.A.S. Hass, P.H.J. Keizers, A. Blok, Y. Hiruma, M. Ubbink, Validation of a lanthanide tag for the analysis of protein dynamics by paramagnetic NMR spectroscopy, *J. Am. Chem. Soc.* 132 (2010) 9952–9953.
- [155] A. Merk, A. Bartesaghi, S. Banerjee, V. Falconieri, P. Rao, M.I. Davis, R. Pragani, M.B. Boxer, L.A. Earl, J.L.S. Milne, S. Subramaniam, Breaking cryo-EM resolution barriers to facilitate drug discovery, *Cell* 165 (2016) 1698–1707.
- [156] K. Mason, N.J. Rogers, E.A. Sutura, I. Kuprov, J.A. Aguilar, A.S. Batsanov, D.S. Yufit, D. Parker, PARASHIFT probes: solution NMR and X-ray structural studies of macrocyclic ytterbium and yttrium complexes, *Inorg. Chem.* 56 (2017) 4028–4038.
- [157] A.C. Harnden, E.A. Sutura, A.S. Batsanov, P.K. Senanayake, M.A. Fox, K. Mason, M. Vonci, E.J.L. McInnes, N.F. Chilton, D. Parker, Unravelling the complexities of pseudocontact shift analysis in lanthanide coordination complexes of differing symmetry, *Angew. Chem. Int. Ed.* 58 (2019) 10290–10294.
- [158] P.H.J. Keizers, B. Mersinli, W. Reinle, J. Donauer, Y. Hiruma, F. Hannemann, M. Overhand, R. Bernhardt, M. Ubbink, A solution model of the complex formed by adrenodoxin and adrenodoxin reductase determined by paramagnetic NMR spectroscopy, *Biochemistry* 49 (2010) 6846–6855.
- [159] L. de la Cruz, T.H.D. Nguyen, K. Ozawa, J. Shin, B. Graham, T. Huber, G. Otting, Binding of low molecular weight inhibitors promotes large conformational changes in the dengue virus NS2B-NS3 protease: fold analysis by pseudocontact shifts, *J. Am. Chem. Soc.* 133 (2011) 19205–19215.
- [160] I. Bertini, V. Calderone, L. Cerofolini, M. Fragai, C.F.G.C. Geraldes, P. Hermann, C. Luchinat, G. Parigi, J.M.C. Teixeira, The catalytic domain of MMP-1 studied through tagged lanthanides, *FEBS Lett.* 586 (2012) 557–567.
- [161] J. Li, K.B. Pilla, Q. Li, Z. Zhang, X. Su, T. Huber, J. Yang, Magic angle spinning NMR structure determination of proteins from pseudocontact shifts, *J. Am. Chem. Soc.* 135 (2013) 8294–8303.
- [162] H. Yagi, Kala B. Pilla, A. Maleckis, B. Graham, T. Huber, G. Otting, Three-dimensional protein fold determination from backbone amide pseudocontact shifts generated by lanthanide tags at multiple sites, *Structure* 21 (2013) 883–890.
- [163] Y. Hiruma, M.A.S. Hass, Y. Kikui, W.-M. Liu, B. Ölmez, S.P. Skinner, A. Blok, A. Kloosterman, H. Koteishi, F. Löhr, H. Schwalbe, M. Nojiri, M. Ubbink, The structure of the cytochrome P450cam–putidaredoxin complex determined by paramagnetic NMR spectroscopy and crystallography, *J. Mol. Biol.* 425 (2013) 4353–4365.
- [164] W.-N. Chen, K.V. Loscha, C. Nitsche, B. Graham, G. Otting, The dengue virus NS2B–NS3 protease retains the closed conformation in the complex with BPTI, *FEBS Lett.* 588 (2014) 2206–2211.
- [165] D. Shishmarev, Y. Wang, C.E. Mason, X.-C. Su, A.J. Oakley, B. Graham, T. Huber, N.E. Dixon, G. Otting, Intramolecular binding mode of the C-terminus of *Escherichia coli* single-stranded DNA binding protein determined by nuclear magnetic resonance spectroscopy, *Nucleic Acids Res.* 42 (2014) 2750–2757.
- [166] L. de la Cruz, W.-N. Chen, B. Graham, G. Otting, Binding mode of the activity-modulating C-terminal segment of NS2B to NS3 in the dengue virus NS2B–NS3 protease, *FEBS J.* 281 (2014) 1517–1533.
- [167] A.R. Camacho-Zarco, F. Munari, M. Wegstroth, W.-M. Liu, M. Ubbink, S. Becker, M. Zweckstetter, Multiple paramagnetic effects through a tagged reporter protein, *Angew. Chem. Int. Ed.* 54 (2015) 336–339.
- [168] D.J. Crick, J.X. Wang, B. Graham, J.D. Swarbrick, H.R. Mott, D. Nietlispach, Integral membrane protein structure determination using pseudocontact shifts, *J. Biomol. NMR* 61 (2015) 197–207.
- [169] Y.-Z. Pan, B. Quade, K.D. Brewer, M. Szabo, J.D. Swarbrick, B. Graham, J. Rizo, Sequence-specific assignment of methyl groups from the neuronal SNARE complex using lanthanide-induced pseudocontact shifts, *J. Biomol. NMR* 66 (2016) 281–293.
- [170] T. Saio, K. Ogura, H. Kumeta, Y. Kobashigawa, K. Shimizu, M. Yokochi, K. Kodama, H. Yamaguchi, H. Tsujishita, F. Inagaki, Ligand-driven conformational changes of MurD visualized by paramagnetic NMR, *Sci. Rep.* 5 (2015) 16685.
- [171] J.-L. Chen, Y. Yang, L.-L. Zhang, H. Liang, T. Huber, X.-C. Su, G. Otting, Analysis of the solution conformations of T4 lysozyme by paramagnetic NMR spectroscopy, *Phys. Chem. Chem. Phys.* 18 (2016) 5850–5859.
- [172] R.S. Ma, Q.F. Li, A.D. Wang, J.H. Zhang, Z.J. Liu, J.H. Wu, X.C. Su, K. Ruan, Determination of pseudocontact shifts of low-populated excited states by NMR chemical exchange saturation transfer, *Phys. Chem. Chem. Phys.* 18 (2016) 13794–13798.
- [173] M. Lescanne, S.P. Skinner, A. Blok, M. Timmer, L. Cerofolini, M. Fragai, C. Luchinat, M. Ubbink, Methyl group assignment using pseudocontact shifts with PARAssign, *J. Biomol. NMR* 69 (2017) 183–195.
- [174] M. Lescanne, P. Ahuja, A. Blok, M. Timmer, T. Akerud, M. Ubbink, Methyl group reorientation under ligand binding probed by pseudocontact shifts, *J. Biomol. NMR* 71 (2018) 275–285.
- [175] D.E. Kamen, S.M. Cahill, M.E. Girvin, Multiple alignment of membrane proteins for measuring residual dipolar couplings using lanthanide ions bound to a small metal chelator, *J. Am. Chem. Soc.* 129 (2007) 1846–1847.
- [176] X. Xu, P.H.J. Keizers, W. Reinle, F. Hannemann, R. Bernhardt, M. Ubbink, Intermolecular dynamics studied by paramagnetic tagging, *J. Biomol. NMR* 43 (2009) 247–254.
- [177] S.C. Chiliveri, J.M. Louis, R. Ghirlando, J.L. Baber, A. Bax, Tilted, uninterrupted, monomeric HIV-1 gp41 transmembrane helix from residual dipolar couplings, *J. Am. Chem. Soc.* 140 (2018) 34–37.
- [178] C.A. Barnes, Y. Shen, J. Ying, Y. Takagi, D.A. Torchia, J.R. Sellers, A. Bax, Remarkable rigidity of the single α -helical domain of myosin-VI As revealed by NMR spectroscopy, *J. Am. Chem. Soc.* 141 (2019) 9004–9017.
- [179] W.-N. Chen, C. Nitsche, K.B. Pilla, B. Graham, T. Huber, C.D. Klein, G. Otting, Sensitive NMR approach for determining the binding mode of tightly binding ligand molecules to protein targets, *J. Am. Chem. Soc.* 138 (2016) 4539–4546.
- [180] D. Xu, B. Li, J. Gao, Z. Liu, X. Niu, G. Nshogoza, J. Zhang, J. Wu, X.-C. Su, W. He, R. Ma, D. Yang, K. Ruan, Ligand proton pseudocontact shifts determined from paramagnetic relaxation dispersion in the limit of NMR intermediate exchange, *J. Phys. Chem. Lett.* 9 (2018) 3361–3367.
- [181] A. Mallagaray, A. Canales, G. Domínguez, J. Jiménez-Barbero, J. Pérez-Castells, A rigid lanthanide binding tag for NMR structural analysis of carbohydrates, *Chem. Commun.* 47 (2011) 7179–7181.
- [182] M. Erdélyi, E. d’Auvergne, A. Navarro-Vázquez, A. Leonov, C. Griesinger, Dynamics of the glycosidic bond: conformational space of lactose, *Chem. Eur. J.* 17 (2011) 9368–9376.
- [183] S. Yamamoto, T. Yamaguchi, M. Erdélyi, C. Griesinger, K. Kato, Paramagnetic lanthanide tagging for NMR conformational analyses of N-linked oligosaccharides, *Chem. Eur. J.* 17 (2011) 9280–9282.
- [184] Y. Zhang, S. Yamamoto, T. Yamaguchi, K. Kato, Application of paramagnetic NMR-validated molecular dynamics simulation to the analysis of a conformational ensemble of a branched oligosaccharide, *Molecules* 17 (2012) 6658–6671.
- [185] A. Gabriel Rabinovich, Diego O. Croci, Regulatory circuits mediated by lectin-glycan interactions in autoimmunity and cancer, *Immunity* 36 (2012) 322–335.

- [186] D.N. Hebert, L. Lamriben, E.T. Powers, J.W. Kelly, The intrinsic and extrinsic effects of N-linked glycans on glycoproteostasis, *Nat. Chem. Biol.* 10 (2014) 902.
- [187] A. Canales, A. Mallagaray, J. Pérez-Castells, I. Boos, C. Unverzagt, S. André, H.-J. Gabius, F.J. Cañada, J. Jiménez-Barbero, Breaking pseudo-symmetry in multiantennary complex N-glycans using lanthanide-binding tags and NMR pseudo-contact shifts, *Angew. Chem. Int. Ed.* 52 (2013) 13789–13793.
- [188] A. Canales, A. Mallagaray, M.Á. Berbís, A. Navarro-Vázquez, G. Domínguez, F.J. Cañada, S. André, H.-J. Gabius, J. Pérez-Castells, J. Jiménez-Barbero, Lanthanide-chelating carbohydrate conjugates are useful tools to characterize carbohydrate conformation in solution and sensitive sensors to detect carbohydrate-protein interactions, *J. Am. Chem. Soc.* 136 (2014) 8011–8017.
- [189] A. Mallagaray, G. Domínguez, T. Peters, J. Pérez-Castells, A rigid lanthanide binding tag to aid NMR studies of a 70 kDa homodimeric coat protein of human norovirus, *Chem. Commun.* 52 (2016) 601–604.
- [190] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, P. Walter, *Molecular Biology of the Cell*, fourth ed., Garland Science, New York, 2002.
- [191] S.S. Wijmenga, B.N.M. van Buuren, The use of NMR methods for conformational studies of nucleic acids, *Prog. Nucl. Magn. Reson. Spectrosc.* 32 (1998) 287–387.
- [192] K. Wüthrich, NMR studies of structure and function of biological macromolecules (nobel lecture), *Angew. Chem. Int. Ed.* 42 (2003) 3340–3363.
- [193] M. Gochin, Nuclear magnetic resonance studies of a paramagnetic metallo DNA complex, *J. Am. Chem. Soc.* 119 (1997) 3377–3378.
- [194] M. Gochin, Nuclear Magnetic resonance characterization of a paramagnetic DNA-drug complex with high spin cobalt; Assignment of the 1H and 31P NMR spectra, and determination of electronic, spectroscopic and molecular properties, *J. Biomol. NMR* 12 (1998) 243–257.
- [195] K. Jaudzems, X. Jia, H. Yagi, D. Zhulenkova, B. Graham, G. Otting, E. Liepinsh, Structural basis for 5'-end-specific recognition of single-stranded DNA by the R3H domain from human Sjbp-2, *J. Mol. Biol.* 424 (2012) 42–53.
- [196] M. Strickland, J. Catazaro, R. Rajasekaran, M.-P. Strub, C. O'Hern, G.A. Bermejo, M.F. Summers, J. Marchant, N. Tjandra, Long-range RNA structural information via a paramagnetically tagged reporter protein, *J. Am. Chem. Soc.* 141 (2019) 1430–1434.
- [197] Y. Song, T.J. Meade, A.V. Astashkin, E.L. Klein, J.H. Enemark, A. Raitsimring, Pulsed dipolar spectroscopy distance measurements in biomacromolecules labeled with Gd(III) markers, *J. Magn. Reson.* 210 (2011) 59–68.
- [198] D. Goldfarb, Gd³⁺ spin labeling for distance measurements by pulse EPR spectroscopy, *Phys. Chem. Chem. Phys.* 16 (2014) 9685–9699.
- DTT*: dithiothreitol
DTTA-C3-yne: diethylene-triamine-tetraacetate propyl-1-yne
DTTA-C4-yne: diethylene-triamine-tetraacetate butyl-1-yne
ECAD12: first two domains of epithelial cadherin
EDTA: ethylenediaminetetraacetic acid
EPR: electron paramagnetic resonance
ERp29: endoplasmic reticulum protein 29
FBDD: fragment-based drug discovery
FM-520: 4-sulphamoyl-N-(4-(trifluoromethyl)benzyl) benzamide
GB1: B1 domain of protein G
GM2 and GM3: second and third discovered monosialic ganglioside
GPS: global positioning system
Grb2 SH2: SH2 domain of growth factor receptor-bound protein 2
HATU: 1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxide hexafluorophosphate
hCA II: human carbonic anhydrase II
HeLa cell: immortal cell line derived from cervical cancer taken from Henriette Lacks
HEPES: 2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulphonic acid
HPLC: high-performance liquid chromatography
HPPK: ATP:2-amino-4-hydroxy-6-hydroxymethyl-7,8-dihydropteridine 6'-diphosphotransferase
Hsp90: heat shock protein 90
HSQC: heteronuclear single quantum correlation
HSQC IPAP: heteronuclear single quantum correlation in-phase/anti-phase
IDA: iminodiacetic acid
IDP: intrinsically disordered protein
IPITG: isopropyl β-D-1-thiogalactopyranoside
LacR: lac repressor
LBP: lanthanide binding peptide
LCT: lanthanide chelating tag
MD: molecular dynamics
MMP-1: matrix metalloproteinase 1
MRI: magnetic resonance imaging
MurD: UDP-N-acetylmuramoylalanine--D-glutamate ligase
NMR: nuclear magnetic resonance
NOE: nuclear Overhauser effect
NOESY: nuclear Overhauser effect spectroscopy
NTA: nitrilotriacetic acid
OB-domain: DNA binding domain of single-stranded DNA binding protein
Paz: domain of the proteins Piwi, Argonaut and Zwiile
PCS: pseudocontact shift
PDZ domain: domain found in post-synaptic density protein, Drosophila disc large tumour suppressor and zonula occludens-1 protein
PDZ protein: family of proteins containing the PDZ domain
PRE: paramagnetic relaxation enhancement
pSRII: phototaxis receptor sensory rhodopsin II
RDC: residual dipolar coupling
RMSD: root mean square deviation
RNA: ribonucleic acid
SAH domain: single α-helical domain
SAP: square antiprism
SH2 domain: Src homology 2 domain
SNAP-25: synaptosomal nerve-associated protein 25
SNARE: SNAP receptor
Src: proto-oncogene tyrosine-protein kinase (Src derived from sarcoma)
SrtA: sortase A
SSBP: single-stranded DNA binding protein
Syt1: synaptotagmin 1
Sμbp2-R3H: R3H domain of Sμbp2
T4 Lys: T4 lysozyme
TAHA: triaminohexaacetate
TCEP: 3,3',3''-phosphanetriyltriopropanoic acid
Tris: 2-amino-2-(hydroxymethyl)propane-1,3-diol
TSAP: twisted square antiprism
U1A: small nuclear ribonucleoprotein polypeptide A
UBA1: ubiquitin-like modifier activating enzyme 1
Ubi: ubiquitin

Glossary

3MDPA: 3-mercapto-2,6-pyridinedicarboxylic acid
4MDPA: 4-mercapto-2,6-pyridinedicarboxylic acid
4MMDPA: 4-mercaptomethyl-dipicolinic acid
4MTDA: 4'-mercapto-2,2': 6',2''-terpyridine-6,6''-dicarboxylic acid
4PS-DPA: (4-(phenylsulphonyl)-pyridine-2,6-dicarboxylic acid
4PS-PyMTA: 4-phenylsulphonyl-(pyridin-2,6-diyl)bismethylenenitrilo tetrakis (acetic acid)
4VPyMTA: 4-vinyl(pyridine-2,6-diyl)bismethylenenitrilo tetrakis(acetic acid)
7TM: seven-helix transmembrane
ArgN: N-terminal domain of the arginine repressor from *E. coli*
AzF: p-azido-L-phenylalanine
BRM: Brahma (*Drosophila* protein)
BTAA: 2-(4-(bis((1-(tert-butyl)-1H-1,2,3-triazol-4-yl)methyl)amino)methyl)-1H-1,2,3-triazol-1-yl)acetic acid
CEST: chemical exchange saturation transfer
ClANP: caged lanthanide NMR probe
Cryo-EM: cryogenic electron microscopy
CS: contact shift
CSP: chemical shift perturbation
Cyt c: cytochrome c
DEER: double electron-electron resonance
DFE: density functional theory
dGMP: deoxyguanosine 5'-monophosphate
DNA: deoxyribonucleic acid
DOTA: 1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid
DPA: dipicolinic acid
DTPA: diethylenetriaminepentaacetic acid

4.2 Application of paramagnetic lanthanoid chelating tags in NMR spectroscopy and their use for the localization of ligands within biomacromolecules

Reproduced from D. Joss, R. Vogel, K. Zimmermann, D. Häussinger, *Reference Module in Chemistry, Molecular Sciences and Chemical Engineering*, **2020**, doi:10.1016/B978-0-12-409547-2.14848-6 with permission of Elsevier (<https://doi.org/10.1016/B978-0-12-409547-2.14848-6>).

Application of Paramagnetic Lanthanoid Chelating Tags in NMR Spectroscopy and Their Use for the Localization of Ligands Within Biomacromolecules

D Joss, R Vogel, K Zimmermann, and D Häussinger, University of Basel, Basel, Switzerland

© 2020 Elsevier Inc. All rights reserved.

1	Introduction	1
1.1	Paramagnetic Nuclear Magnetic Resonance Spectroscopy	1
2	Lanthanoid Chelating Tags	5
2.1	Introduction and Overview	5
2.2	EDTA-Based Lanthanoid Chelating Tags	5
2.3	DTTA- and DTPA-Based Lanthanoid Chelating Tags	6
2.4	TAHA-Based Lanthanoid Chelating Tags	7
2.5	IDA-, NTA- and Pyridine-Based Lanthanoid Chelating Tags	7
2.6	DOTA-Based Lanthanoid Chelating Tags	8
3	Applications of Lanthanoid Chelating Tags in the Localization of Ligands Within Biomacromolecules	12
3.1	Introduction	12
3.2	Weakly Binding Ligands	12
3.3	Ligands in the Intermediate Exchange Regime	13
3.4	Strongly Binding Ligands	14
4	Localization of Ligands Within Human Carbonic Anhydrase II Using ¹⁹F Pseudocontact Shift Analysis	15
4.1	Human Carbonic Anhydrase and Its Use in the Development of Artificial Metalloenzymes	15
4.2	Localization of Ligands within Human Carbonic Anhydrase II—Introduction and Concept	16
4.3	PCS Analysis and Determination of the Anisotropy Parameters from Multiple Positions Within the Protein	18
4.4	Measurement of ¹⁹ F PCS and Localization of the Ligand Using a Monte-Carlo Approach	19
4.5	Determination of the Ligand Position by Analysis of the Isosurfaces of the Anisotropy Parameters	24
4.6	¹ H- ¹⁹ F-HOESY Experiments for Determination of ¹ H PCS of the Ligand and Protein–Ligand Interactions	24
5	Conclusion and Outlook	26
	Acknowledgment	26
	References	26

1 Introduction

1.1 Paramagnetic Nuclear Magnetic Resonance Spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy is ideally suited for the elucidation of structure, dynamics and interactions of biomacromolecules in solution.¹ The development of homo- and heteronuclear experiments in order to establish through-bond^{2–5} and through-space correlations^{6,7} allow for the complete assignment of the protein by experiments linking the amide proton with the backbone and side chain spins from the obtained through-bond correlations. The three-dimensional structure of the protein is then computed using the structural restraints obtained (Fig. 1).^{6,7,10,11} The possibility to tune and adjust the sample conditions, e.g. buffer composition, salt concentration, temperature, additives and pH, allows for realistic experiments that can be adjusted to specific physiological conditions.¹ Furthermore, the dynamics of the protein of interest can be investigated by analysis of relaxation times and heteronuclear Overhauser effects (hetNOEs).^{12,13} Besides the convenient application of advanced NMR techniques to proteins, assignment techniques and further experiments were developed for nucleic acids in order to elucidate their structure, dynamics and protein–nucleic acid complexes.^{14–17} However, the distance limitation of the exploited nuclear Overhauser effect (NOE) of 5–8 Å for structure calculations and unspecific chemical shift perturbation (CSP) experiments, used for the monitoring of interactions of biomacromolecules and ligand-binding, render more advanced methods that yield structural restraints over larger distances highly desirable. Paramagnetic NMR can contribute in this regard by providing long-range, through-space interactions in order to obtain distance and orientational restraints of biomacromolecules and their complexes with other biomacromolecules and ligand molecules.¹⁸

An impressive and early study described the quantitative determination of the conformations of mononucleotides in solution using lanthanoid induced shifts,¹⁹ and in the following, Bertini, Banci, Luchinat et al. exploited the paramagnetism of lanthanoids by substitution of diamagnetic metal ions in metalloproteins and demonstrated thereby thoroughly the application and use of paramagnetic NMR in structural biology.^{20–23} In order to enable the introduction of lanthanoid centers into proteins without native metal site in order to take advantage of the paramagnetism-based structural restraints, lanthanoid binding peptides (LBPs) that imitate the zinc finger and EF hand motif of native metalloproteins were developed^{24–27} and can be added to the N- or C-terminus of the protein of interest. In order to provide more rigid chelators and attach an LCT at any possible position within the scaffold of the biomacromolecule, a large variety of double- and single-armed lanthanoid chelating tags (LCTs) has been developed.^{28–37} LCTs have since then found widespread applications in structural biology and biomolecular NMR, e.g. structural characterization of

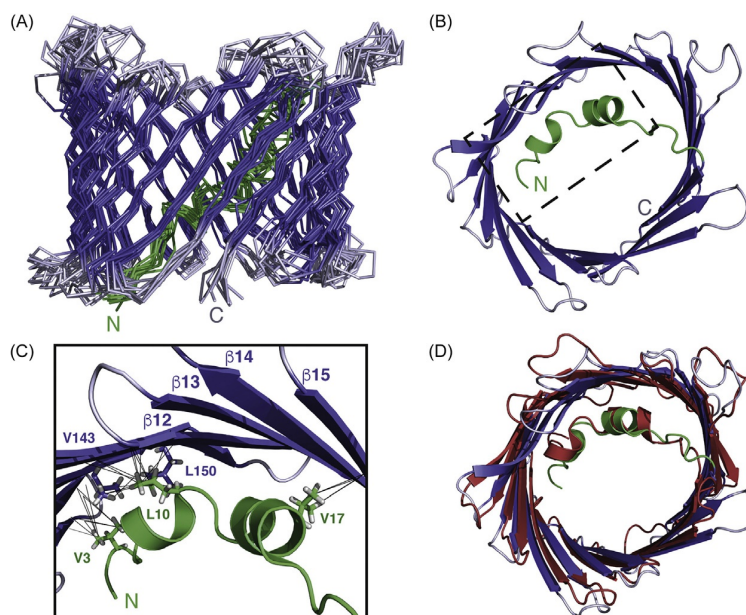


Fig. 1 Solution NMR Structure of hVDAC-1 in LDAO Detergent Micelles refined using NMR restraints. Backbone representation of the 10 conformers with lowest violation of dihedral angles and NOE restraints (A, N-terminus: green, loops: light blue, β -strands: blue). Refined conformer closest to the mean (B, top view). NOE contacts between the N-terminal helices and the wall of the membrane protein (C, NOE contacts: black, N-terminus: green, loops: light blue, β -strands: blue). Overlay of the refined solution NMR structure of hVDAC-1 with the crystal structure of mVDAC-1 (D, PDB 3EMN⁸).⁹ NMR Studies of Structure and Function of Biological Macromolecules (Nobel Lecture). Wüthrich, K.; *Angew. Chem. Int. Ed. Engl.* **2003**, *42*, 3340–3363.

antibiotic drug targets (Fig. 2),³⁹ localization of ligands within proteins,^{40,41} observation of long-range RNA structural information using a paramagnetically tagged reporter protein,⁴² enhancement of signal dispersion in studies of intrinsically disordered proteins,⁴³ structural analysis of biologically highly relevant, large complexes of biomacromolecules⁴⁴ and characterization of the interactions between glycans and proteins.⁴⁵

The introduction of paramagnetic centers in biomacromolecules by either direct complexation by metalloproteins, modification of the protein construct with an LBP or attachment of a small-molecule chelator with appropriate linker moiety leads to sizeable effects visible in NMR experiments that were thoroughly and critically reviewed in multiple reviews.^{18,46–64} The most important effects associated with induced anisotropy caused by paramagnetic centers comprise pseudocontact shifts (PCS),^{28,30,33,34,36,37,65–72} residual dipolar couplings (RDC)^{73–77} and paramagnetic relaxation enhancement (PRE).^{78–84} PCSs and RDCs are exemplified in Fig. 3 by depicting a concrete example of an in-cell NMR study, that delivered *de-novo* derived protein structures from 2D NMR experiments. Besides the most important paramagnetic effects as PCS, RDC and PRE, also contact shifts (CS),^{85–88} residual anisotropic chemical shifts (RACS),^{89,90} dipolar shift anisotropy (DSA) as well as cross-correlation between DSA and dipolar relaxation (DSA/DD)⁹¹ are associated with the incorporation of a paramagnetic center in a target of interest.

Most importantly, the shifts caused by through-space PCS effects provide angular and distance information about the nucleus. Upon application of PCS NMR on proteins, the sequential assignment of backbone amides can be combined with PCS shift data and, therefore, be used in order to obtain structural restraints for each backbone N and H spin. Due to their distance dependence of

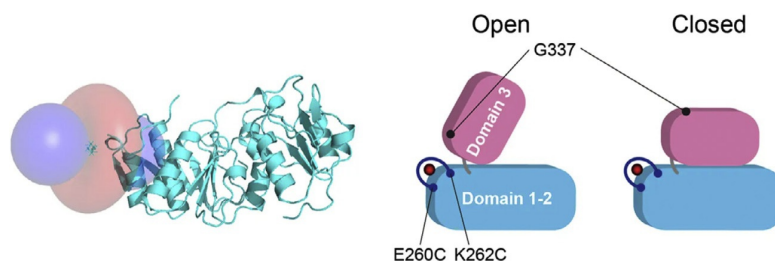


Fig. 2 Caged lanthanoid NMR probe (CLaNP-5)²⁸ attached to a double cysteine mutant of MurD for monitoring of conformational changes. Induced isosurfaces of the Yb³⁺-loaded LCT when ligated to domains 1 and 2 (PCS +/- 0.5 ppm, blue: positive, red: negative, PDB: 3UAG,³⁸ left). Monitoring of the open and closed conformation of domain 3 using PCS restraints (right).³⁹ Ligand-driven conformational changes of MurD visualized by paramagnetic NMR. Saio, T.; Ogura, K.; Kumeta, H.; Kobashigawa, Y.; Shimizu, K.; Yokochi, M.; Kodama, K.; Yamaguchi, H.; Tsujishita, H.; Inagaki, F. *Sci. Rep.* **2015**, *5*, 16685. Nature Publishing Group.

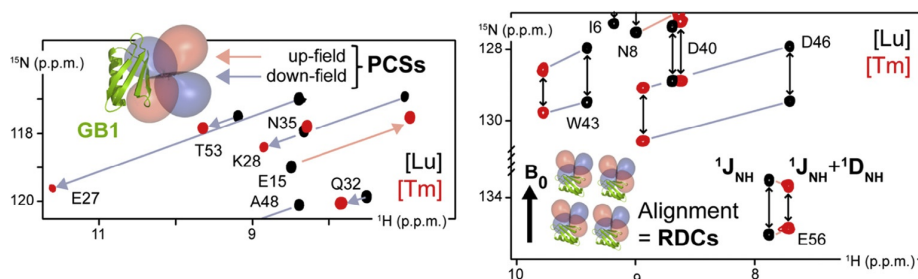


Fig. 3 PCS induced by a Tm-DOTA-M7Py in comparison with its diamagnetic reference Lu-DOTA-M7Py observed on a GB1 E19C construct in ^1H - ^{15}N heteronuclear single quantum coherence (HSQC) experiments (PCS \pm 2.0 ppm, blue: positive, red: negative, *left*). Partial alignment of the labeled protein to the external magnetic field B_0 leading to RDCs observed in ^1H - ^{15}N HSQC IPAP experiments.³³ Reproduced with permission from Müntener, T.; Häussinger, D.; Selenko, P.; Theillet, F.-X. *J. Phys. Chem. Lett.* **2016**, *7*, 2821–2825. Copyright 2016 American Chemical Society.

R^{-3} , PCSs can yield structural restraints over distances of more than 200 Å in favorable cases.¹⁸ In contrast to the through space nature of PCSs, CSs are through-bond effects and only transferred over a few chemical bonds and are, therefore, less interesting for the analysis of biomacromolecules. However, CSs can give valuable insights into ligand structure and conformation, since the observed shifts are dependent on the structure and bonding interactions.^{92,93}

RDCs are especially well suited for structural analysis of large biomacromolecules, since the effect is independent of the distance between the paramagnetic center and the measured nucleus and yields the orientation of the N—H bond vectors in space as structural restraints. RDCs can either be induced by partial alignment of the biomacromolecules by incorporation of a paramagnetic lanthanoid ion with anisotropic electron distribution or by application of an alignment medium as bacteriophages, acrylamide gels or bicelles.^{73,77,94} The RDCs induced by incorporation of paramagnetic centers can be diminished due to internal flexibility of the biomacromolecule, however, this effect can then be further exploited to determine, for example, the stiffness of an α -helix and to draw conclusions about the biological significance of the gained insight.⁷⁷ Combination of LCT-induced PCSs and RDCs was used to determine backbone structures of biomacromolecules inside intact eukaryotic cells, as a unique property of covalently bound high affinity LCTs is their compatibility with *in-cellulo* experiments.^{33,69,95}

PRE is, in a first approximation, an isotropic effect and thus yields only distance and no angular information. Furthermore, the magnitude of the effect decreases with R^{-6} with respect to the distance between the observed spin and the paramagnetic center. However, PREs are not very susceptible to motional averaging, are convenient to induce by commercially available spin labels and are more straightforward to analyze than PCSs and RDCs. Nonetheless, since PCSs provide both distance and angular information, exhibit a distance dependence of R^{-3} and can be observed on any NMR active nuclei, e.g. ^1H , ^{13}C , ^{15}N , ^{19}F , or ^{31}P by one- or multidimensional NMR experiments, PCSs are perfectly well suited for the localization of ligands within biomacromolecules.

The PCS is a dipolar, through-space effect arising from the dipolar interaction between the nuclear spin and, in the case of LCTs, the orbital momentum of the unpaired f-electrons of the paramagnetic lanthanoid center that manifests in NMR experiments by a shift contribution to the NMR resonance of the detected nucleus, depending on the positioning of the nuclear spin with respect to the metal center.⁹⁶ Since the distance between the spin and the paramagnetic metal center is much larger than the ionic radius of the metal center, the point-dipole approximation can be applied that depicts the magnetic dipole moment as located in a single point.⁶² The PCS can then be translated into structural restraints by fitting of the experimental data to a known or calculated structure using software packages like Numbat⁹⁷ or Paramagpy.⁹⁸ Programs capable of PCS analysis minimize the difference between observed and back-calculated PCS in an iterative fashion until the convergence is achieved. The resulting axial and rhombic parameters of the observed anisotropy tensors should be reported in the unique tensor representation (UTR) using the zyz' convention, as only this representation allows for direct comparison with tensors from other LCTs.⁶² In order to visualize the anisotropy parameters induced by the lanthanoid within the biomacromolecule in a suitable fashion, isosurfaces for given ppm values are displayed graphically (Fig. 4).

Thereby, the strength of the shifts in different regions of a biomacromolecule can be directly and visually inspected, e.g. with the program PyMOL. While red parts by definition indicate regions with negative shifts, blue parts constitute regions with positively shifted NMR signals. By the use of PCS analysis, each N and H nucleus can be localized in the three-dimensional space, and therefore valuable structural information can be gained about the structure of the protein of interest. More specifically, a *de-novo* derived structure can be obtained by starting the refinement process from an *in silico* structure, e.g. by use of the Rosetta database. Structures from X-ray crystallography can be refined in solution by PCS and crucial differences between the solution- and solid-state structures, e.g. a different ordering or conformation of more flexible regions within a protein or the alignment of different protein domains, can immediately be detected due to poor fits of the data set obtained via paramagnetic NMR spectroscopy.^{39,99} For nucleic acids, the methodology can be applied e.g. with total correlation spectroscopy (TOCSY) or nuclear Overhauser effect spectroscopy (NOESY) spectra.⁷² The PCS can be calculated by use of the following equation that can be displayed both in polar or Cartesian coordinates⁹⁶:

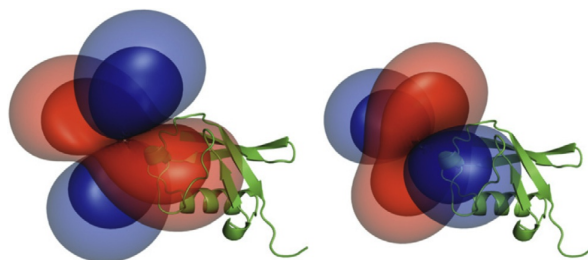


Fig. 4 Isosurfaces induced by the attachment of Ln-P4T-DOTA complexes (Dy^{3+} : left, Tm^{3+} : right) and their relative orientation to ubiquitin S57C (PCS isosurfaces: 1.5 ppm (outer layer), 4.0 ppm (inner layer)).³⁷ Reproduced by permission of The Royal Society of Chemistry from Joss, D.; Häußinger, D. *Chem. Commun.* **2019**, 55, 10543–10546.

$$\text{PCS (polar coordinates)} \delta^{PCS} = \frac{1}{12\pi r^3} \left[\Delta\chi_{ax} (3 \cos^2 \theta - 1) + \frac{3}{2} \Delta\chi_{rh} \sin^2 \theta \cos 2\phi \right] \quad (1)$$

$$\text{PCS (Cartesian coordinates)} \delta^{PCS}(x, y, z) = \frac{1}{12\pi r^3} \left[\Delta\chi_{ax} \frac{2z^2 - x^2 - y^2}{r^2} + \frac{3}{2} \Delta\chi_{rh} \frac{x^2 - y^2}{r^2} \right] \quad (2)$$

For the complete description of the anisotropic part of the magnetic susceptibility that is induced by the lanthanoid ion incorporated in or attached to a biomacromolecule of interest, in total eight parameters are required: axial and rhombic component of the magnetic susceptibility tensor ($\Delta\chi_{ax}$ and $\Delta\chi_{rh}$), the coordinates of the paramagnetic center (x, y, z) and the Euler angles describing the orientation with regard to the coordinate frame of the protein (α, β, γ). The anisotropy parameters $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$ describe the directional asymmetry of a tensor using the following definitions⁹⁶:

$$\text{Axiality } \Delta\chi_{ax} = \chi_{zz} - \frac{\chi_{xx} + \chi_{yy}}{2} \quad (3)$$

$$\text{Rhombicity } \Delta\chi_{rh} = \chi_{xx} - \chi_{yy} \quad (4)$$

As discussed, software packages, e.g. Numbat⁹⁷ or Paramagpy⁹⁸ are used to derive the anisotropy parameters from the experimental data by minimizing the difference between the experimental and back-calculated PCS for a given structure (Fig. 5).

Alternative approaches use a given set of PCS in combination with a known or partially known structure for assignment of methyl groups^{100,101} or the localization of ligand atoms^{40,41,102} within a biomacromolecule of interest.

The aim of this case study is to provide the reader with basic knowledge about LCTs and the application of LCTs for the localization of ligands within biomacromolecules using PCS analysis with a particular focus on the localization of fluorinated ligands within human carbonic anhydrase II (hCA II) using PCS as long-range restraints with a high accuracy over long metal-fluorine distances.

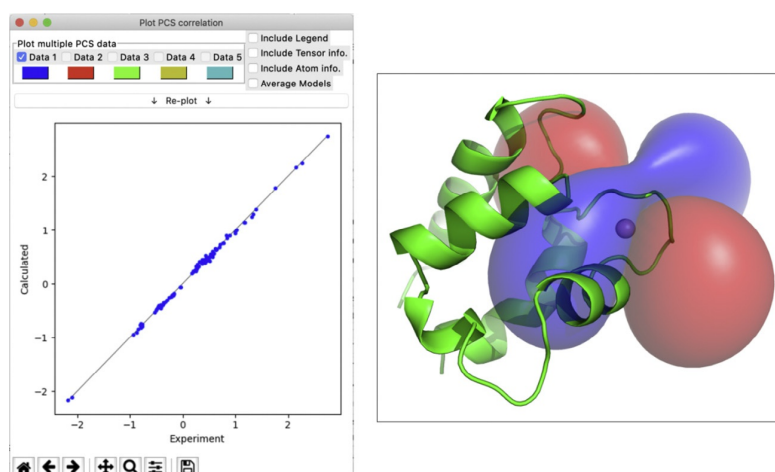


Fig. 5 View on a part of the interface of the novel software package Paramagpy for PCS fitting. Correlation plot of experimental and calculated PCS (left) and isosurfaces of Er^{3+} incorporated in calbindin D9k (PyMOL, right).⁹⁸ Paramagpy: Software for Fitting Magnetic Susceptibility Tensors Using Paramagnetic Effects Measured in NMR Spectra. Orton, H. W.; Huber, T.; Otting, G. *Magn. Reson. Discuss.* **2019**, 1–18. Copernicus publications

2 Lanthanoid Chelating Tags

2.1 Introduction and Overview

In order to exploit the paramagnetic effects induced by lanthanoids in biomacromolecules, the most important and basic consideration is the incorporation of the lanthanoid center into the biological target of interest. As demonstrated by Bertini et al., the most simple and convenient starting point is a protein that possesses a native metal binding site, e.g. calbindin D9k. Due to the similar oxophilicity and ionic radius of the members of the lanthanoid series when compared with Ca^{2+} ions,¹⁰³ paramagnetic lanthanoid ions can often be incorporated in calcium binding sites. Given that the protein exhibits only one calcium binding site or only one calcium binding site that is amenable for substitution with a lanthanoid ion, the induced paramagnetic effects can then be used as structural restraints for the characterization of the protein scaffold.¹⁰⁴ Furthermore, the fixed protein coordination site can also be used for a study of the induced anisotropy parameters by the full series of lanthanoids.^{47,105} However, it is important to note that the displayed anisotropy of lanthanoid centers can depend strongly on the type of donor atoms and the coordination polyhedron of the lanthanoid complex.^{106–110} While the direct substitution of calcium by lanthanoid ions is highly interesting for any protein containing a suitable metal binding site, the incorporation of lanthanoid ions into other proteins has to rely on external tags that are fused to the protein during expression or by attachment to specific residues on the surface of the protein.

LBP mimics the native metal binding site of metalloproteins and can conveniently be inserted into the sequence or fused to the termini of the protein.^{25–27,111,112} In order to enhance the rigidity of the lanthanoid binding site delivered by the LBP with respect to the protein scaffold, two-point anchored LBPs, i.e. an LBP linked to the protein through one peptide fusion and one disulfide bond, present an appealing improvement.²⁷ Although this modification always leads to complete ligation yields for the attachment of the tag to the protein, significant drawbacks are associated with LBPs, i.e. the incorporation of the tag is mostly performed at the termini of the sequence since otherwise a significant distortion of the protein scaffold is possible, the additional residues lead to more crowded protein NMR spectra and the rigidity as well as properties of the chelator cannot be that specifically tuned as e.g. with chelators based on ethylenediaminetetraacetic acid (EDTA) or 1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid (DOTA) scaffolds. To avoid these drawbacks, LBPs that circumvent this severe restriction by anchoring the LBP via cysteine residues to the protein were proposed.^{113,114}

In order to enable the tuning of the induced paramagnetic effects and the incorporation of different linker moieties and ligand donor atoms, the attachment of common metal chelators as EDTA,^{115,116} diethylenetriaminepentaacetic acid (DTPA),¹¹⁷ or DOTA^{28,29} and their site-specific attachment to single cysteine residues proved to be highly successful (Fig. 6). Cysteine residues are specifically well-suited to serve as attachment point of an LCT, since they can be introduced on the surface of the protein by mutation by replacing a serine for a cysteine residue without distortion of the local structure. Most importantly, the nucleophilicity of cysteines, due to the significant fraction of deprotonated cysteine thiols around the physiological pH, renders them very attractive for ligation to an LCT with an electrophilic attachment point. Furthermore, cysteines display a low abundance of only 1.4% in protein sequences (average of proteins contained in the UniProtKB/Swiss-Prot data bank).¹¹⁸

Synthetic LCTs can be designed in virtually any substitution and donor atom pattern¹⁸ and allow a high variety of thiol-reactive moieties to be incorporated, i.e. activated mixed disulfides,^{29,34,36} activators with pyridinesulfone leaving groups,^{33,69,119} pyridine-thiazole moieties,^{35,37} bioorthogonal activators,^{67,120} or two-point anchoring disulfides that strongly restrict the flexibility of the LCT on the surface of the protein.^{28,121–124}

However, a serine to cysteine mutation can be problematic for proteins that show instability upon displaying cysteine residues on their surface or for proteins that display multiple cysteine residues, which are required in order to fulfill their biological function, on their surface. In the future, bio-orthogonal conjugation to artificial amino acids that can be incorporated into the genetic code, are expected to play a key role in the incorporation of LCTs into proteins and paramagnetic NMR spectroscopy of biomacromolecules.

2.2 EDTA-Based Lanthanoid Chelating Tags

In order to evolve the methodology of inducing paramagnetic effects by lanthanoid-loaded peptide motifs further to a more versatile chelator that is independent of the process of protein expression, suitable synthetic chelators were required and auspicious

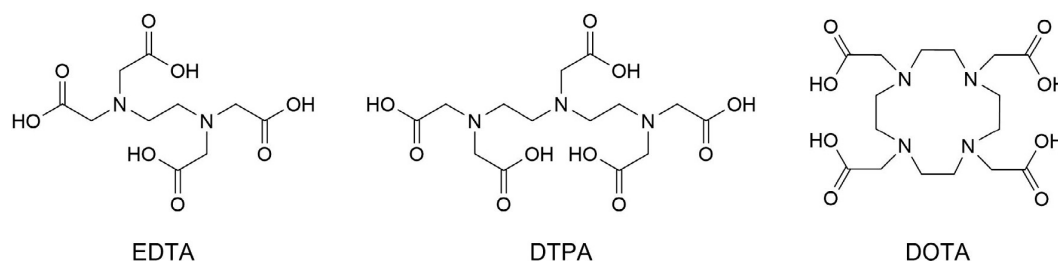


Fig. 6 Scaffolds of chelators used for the synthesis of LCTs.

candidates were EDTA and DTPA scaffolds. The development of LCTs started therefore with the synthesis of EDTA-derived LCTs that were then conjugated to a cysteine residue of a protein in order to observe structural restraints on covalently tagged proteins.

The development of an EDTA-derived LCT and its application for paramagnetic protein NMR was first reported by Dvoretzky et al.¹²⁵ The Yb³⁺-loaded chelator equipped with a pyridine-2-thiol activated linker moiety was attached to barnase H102C and the obtained structural restraints, i.e. PCS and RDC, were fitted to the X-ray structure of wild-type barnase.

Two years later, an enantiomerically pure EDTA-based LCT with a disulfide linker was proposed by Ikegami et al.¹¹⁵ Upon conjugation of the LCT to trigger factor, a molecular chaperone associated with the ribosome, RDCs of up to 8 Hz were observed at 800 MHz proton frequency. However, upon equipping the EDTA-based LCT with a linker, and coordination of the metal that leads to a fixed geometry, a nitrogen atom within the chelator becomes a stereogenic center and two diastereomeric species are formed. The two diastereomeric species lead, when attached to the protein of interest, to a second set of peaks in ¹H-¹⁵N HSQC spectra. A significant improvement was then achieved by Leonov et al., who attached the linker of the LCT to the carbon backbone of the EDTA ligand instead of the nitrogen atom.¹²⁶ Thereby, each of the two nitrogen atoms is substituted by two identical groups and despite the fact, that still two enantiomers of the LCT can be formed, in favorable cases only one set of signals was observed in ¹H-¹⁵N HSQC experiments.

In 2006, two novel EDTA chelators were reported by Haberz et al. that lead to linearly independent anisotropy parameters when compared to previously synthesized EDTA-tags.¹²⁷ Interestingly, the LCTs were applied for the solution structure refinement of trigger factor and the authors demonstrated that the affinity of the newly synthesized LCTs for lanthanoids is higher than the one of the native metal binding site in calmodulin. By the design and synthesis of EDTA-derived LCTs, it was successfully demonstrated that the methodology of the LBPs can be successfully transferred to the use of synthetic metal chelators. The important advantage of the newly synthesized LCTs is that they can be attached at multiple sites within a protein. However, the observed anisotropy parameters were only of modest size and for some of the EDTA-based tags peak doubling was observed that severely complicates assignment of the spectra. Therefore, other metal chelators were tested in order to serve as scaffolds for LCTs, as e.g. diethylene-triaminetetraacetic acid (DTTA) and DTPA ligands.

2.3 DTTA- and DTPA-Based Lanthanoid Chelating Tags

In order to improve the initially developed EDTA tags, DTTA- and DTPA-derived LCTs were synthesized and applied in paramagnetic NMR of biomacromolecules.

As a first example, Prudêncio et al. proposed in 2004 a DTPA-based LCT with two linker moieties that was termed CLaNP and is attached to suitable double-cysteine mutants of the protein of interest.¹¹⁷ However, the authors detected up to five signals per residue in ¹H-¹⁵N HSQC spectra. In Ln-DTPA-bisamide complexes, in theory up to eight different stereoisomers can be formed, although it is likely that the occurrence of some species is promoted by interaction of the chelator with the surface of the protein.^{62,128} Although the occurrence of multiple species was rather discouraging, observable PCSs were detected even on nuclei further away than 40 Å, a result that confirmed that valuable long-range restraints can be obtained by using synthetic paramagnetic metal tags. Upon investigation of DTPA-based LCTs that lead to multiple detected species, DTTA-derived LCTs were envisioned to be synthesized.

Jiang et al. developed bio-orthogonally linked diethylene-triamine-tetraacetate propyl-1-yne (DTTA-C3-yne) and diethylene-triamine-tetraacetate butyl-1-yne (DTTA-C4-yne) LCTs that are attached to the protein of interest by the use of a copper-catalyzed azide alkyne click reaction of an alkyne moiety on the LCT with a *p*-azido-*L*-phenylalanine (AzF) residue of the protein (Fig. 7).¹²⁰ Thereby, the authors demonstrated that large anisotropy parameters can be obtained and RDCs up to 8 Hz were detected using Tb³⁺ and Tm³⁺ loaded LCTs on ubiquitin and enzyme II B. Based on intermolecular PCS measured for a UBA1-ubiquitin complex, the authors reported a refined model on the interactions between the two proteins.

Two further DTTA-derived LCTs equipped with pyridinesulfone linker moieties were reported in 2017 by Chen et al.¹²⁹ Both the 4PS-PyDTTA and the methyl-substituted analogue 4PS-6 M-PyDTTA were tested on ubiquitin and SrtA constructs and yielded

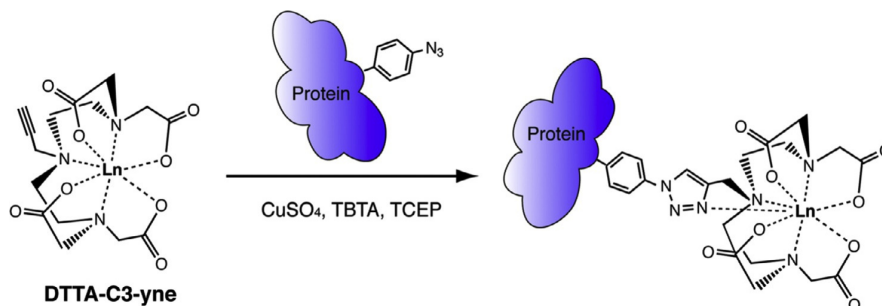


Fig. 7 Copper catalyzed Huisgen azide alkyne cycloaddition of DTTA-C3-yne to a 4-azidophenylalanine residue of the target protein. Remarkably, the triazole ring forms during the ligation reaction a further coordinating nitrogen site and thereby creates a very short linkage from the LCT to the protein of interest.¹²⁰ Lanthanoid tagging via an unnatural amino acid for protein structure characterization. Jiang, W.-X.; Gu, X.-H.; Dong, X.; Tang, C. J. *Biomol. NMR* **2017**, *67*, 273–282. Springer.

large PCSs and RDCs. In line with the results of Haberz et al. for the EDTA-derived LCT,¹²⁷ the authors of the present study observed a high stability of the nonmethylated lanthanoid complex, i.e. the affinity of the LCT to the lanthanoid ion exceeds the one of the native calcium binding site in SrtA. Interestingly, 4PS-6 M-PyDTTA was more susceptible to the addition of EDTA than its nonmethylated analogue.¹²⁹ However, the methyl substitution leads to a significant increase in the anisotropy parameters (4PS-6 M-PyDTTA: $\Delta\chi_{ax} = 11.0 \times 10^{-32} \text{ m}^3$ and $\Delta\chi_{rh} = 3.5 \times 10^{-32} \text{ m}^3$ vs. 4PS-PyDTTA: $\Delta\chi_{ax} = 7.6 \times 10^{-32} \text{ m}^3$ and $\Delta\chi_{rh} = 3.0 \times 10^{-32} \text{ m}^3$ for Tm^{3+} on SrtA D82C), which suggests that the translational and rotational averaging of the paramagnetic effects by reorientation of the LCT on the surface of the protein is suppressed.

While the DTPA-derived LCTs can lead in theory to observation of up to eight signal sets for each amide in ^1H - ^{15}N HSQC spectra and in practice up to five different signal sets were reported, the DTTA-based LCTs equipped with bio-orthogonal and pyridine-sulfone linker provide one set of signals and yield sizeable anisotropy parameters. Nonetheless, further improvement of the anisotropy parameters and increase of stability under all buffer conditions remains an interesting target.

2.4 TAHA-Based Lanthanoid Chelating Tags

In order to investigate the use of further chelators for the use as LCT in paramagnetic NMR spectroscopy, Peters et al. demonstrated in 2011 that cysteinyl-phenyl-triaminohexaacetate (Cys-Ph-TAHA) can be used as suitable scaffold for the design and synthesis of an LCT (Fig. 8).³⁰ Advantageously, Cys-Ph-TAHA is highly symmetric, leads therefore to only one observed set of signals in ^1H - ^{15}N HSQC spectra and shows thermal stability over a wide range of temperature (278–315 K). Furthermore, a metal affinity in the femtomolar range is provided and PCSs up to 2 ppm and RDCs up to 18 Hz are induced by the novel LCT. The authors of the study conjugated Cys-Ph-TAHA to a 90 kDa complex of Lac repressor protein, deoxyribonucleic acid (DNA) and isopropyl β -D-1-thiogalactopyranoside (IPTG). A minor disadvantage of the presented LCT is its pH-sensitivity that excludes the use of the tag for studies at pH values below 4.0, e.g. in order to monitor the unfolding process of a protein of interest under acidic conditions.

2.5 IDA-, NTA- and Pyridine-Based Lanthanoid Chelating Tags

Simple, acyclic chelators captivate the interest of researchers due to their cheap and easily accessible precursors. Although it is clear that the induced anisotropy parameters are expected to be smaller than with extremely rigid, sterically crowded, macrocyclic LCTs, simple acyclic chelators can constitute an interesting alternative to more advanced LCTs for applications that involve only small- and medium-sized proteins. Small LCTs are conveniently built from picolinic acid derivatives or iminodiacetic acid (IDA) and nitrilotriacetic acid (NTA)-scaffolds that are closely related to their larger analogues DTTA and DTPA.¹⁸

Su et al. developed a 4-mercaptomethyl-dipicolinic acid (4MMDPA) LCT that yields PCS up to 2 ppm for the Yb^{3+} -complex when conjugated to the N-terminal domain of arginine repressor protein (ArgN).⁶⁵ Upon attachment of the LCT to a cysteine residue in position 68 of ArgN by using Ellman's reagent, the LCT is further restricted in its mobility relative to the protein surface via the coordination of a nearby carboxylate of the protein scaffold to the lanthanoid. Closely related to the LCT presented by Su et al.,⁶⁵ Man et al. reported a 3-mercapto-2,6-pyridinedicarboxylic acid (3MDPA) LCT that bears a shorter linker in the 3-position of the pyridine instead of the 4-position.¹³⁰ Upon conjugation of the LCT to ArgN and T4 lysozyme C54T C97A Q69C, the LCT was shown to yield for such a small-molecule chelator sizeable anisotropy parameters for the Tm^{3+} loaded construct. As in the case of 4MMDPA, the LCT is suspected to interact with a carboxylic acid group of the Glu21 sidechain of the protein. Jia et al. extended the repertoire of dipicolinic acid derived LCTs by the design and synthesis of 4MDPA that includes a short linker in the 4-position of the pyridine scaffold.¹³¹ While the tag offers quantitative tagging yields, the obtained anisotropy parameters still offer the potential of significant advances. In order to enable the measurement of PCS and RDC effect under reductive conditions as e.g. found while performing in-cell NMR studies, it would be highly desirable to develop a reduction-stable linker moiety based on the dipicolinic acid LCTs.

Li et al. developed therefore in 2011 a dipicolinic acid-based LCT, which bears a vinylpyridine moiety that reacts in a thiol-ene reaction with the cysteine thiol of the protein, and was demonstrated to be applicable on both ArgN as well as ubiquitin (Fig. 9).¹³² The protein-LCT conjugate was shown to be stable both in dithiothreitol (DTT) and 3,3',3''-phosphanetriyltripropanoic acid (TCEP) due to the reductively stable thioether that is formed during the tagging reaction. The LCT induces sizeable effects both on ArgN and ubiquitin constructs.

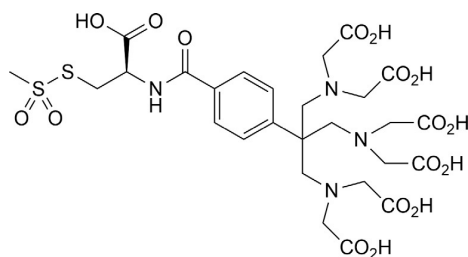


Fig. 8 Chemical structure of the Cys-Ph-TAHA LCT³⁰ synthesized in a total yield of 28% over seven steps.

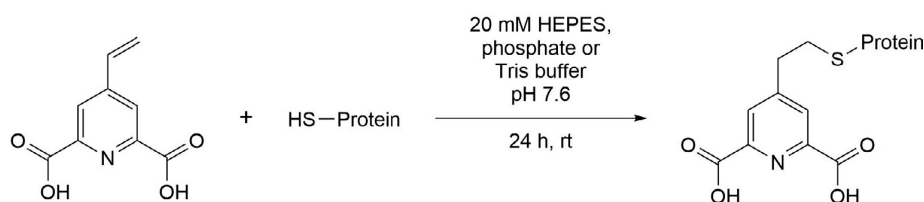


Fig. 9 Nucleophilic addition of the cysteine residue of a protein via thiol-ene reaction to the LCT reported by Li et al.¹³² in aqueous solution in order to create a reduction-stable linkage.

Since the interaction of the small molecule chelators with sidechain carboxylates can directly be used in order to immobilize the LCT on the surface, Swarbrick et al. developed a surprisingly simple LCT based on IDA.¹³³ Upon conjugation to the protein on a residue in an α -helix via a disulfide-bond, the LCT was shown to interact with an Asp residue in the $i + 4$ position by its free coordination sites. The synthesis of the LCT shows an overall yield of 45% for three steps, a result that demonstrates the straightforward availability of this LCT. However, the presented LCT is not generally applicable, since a basic residue in the $i + 4$ position, which is either present in the native state of the protein or has to be introduced by mutation, is required. Following the study with IDA-SH tags, Swarbrick et al. demonstrated that instead of IDA-based tags, also NTA-derived LCTs can be used for the generation of structural restraints using immobilization by a nearby carboxylate or with a second NTA-LCT in the $i + 4$ position resulting in an octadentate coordination motif.¹³⁴ In order to obtain large structural restraints by additional coordination to the IDA-SH and NTA-SH LCTs, Yagi et al. determined in a comprehensive study the optimal position of coordinating carboxylate groups of the sidechain in relation to the tag position.¹³⁵ While IDA-SH tags require additional coordination in order to yield large PCS, NTA-SH LCTs can also yield large paramagnetic effects independent of additional coordination by carboxylate groups of the protein.

Interestingly, Loh et al. incorporated an alkyne moiety in both NTA- and IDA-derived LCTs, giving rise to the possibility of bio-orthogonal tagging reactions, i.e. a copper catalyzed click reaction between the LCT equipped with an alkyne and an azide functional group of an AzF residue incorporated in the protein.¹³⁶ Thereby, the authors obtained large anisotropy parameters on a GB1 V21 construct using the alkyne-NTA LCT ($\Delta\chi_{ax} = 16.3 \times 10^{-32} \text{ m}^3$ and $\Delta\chi_{rh} = 7.8 \times 10^{-32} \text{ m}^3$ for Tm^{3+}) and demonstrated the applicability of the novel LCT by conjugation to AzF-containing constructs of PpiB and ERp29. LCTs that offer the possibility for bio-orthogonal tagging reactions are generally applicable, even to proteins, that show structural distortions upon introduction of an additional cysteine or lose their biological activity.

Taking advantage both from vinylpyridine LCTs as well as IDA- and NTA-derived LCTs led to the synthesis of the 4-vinyl(pyridine-2,6-diyl)bismethylenetrinitrilo tetrakis(acetic acid) (4VPyMTA) LCT by Yang et al.¹³⁷ Most likely due to the increased flexibility of the chelator and its pendant arms as well as the flexible linker moiety that can lead to averaging of paramagnetic effects, the newly developed LCT displays rather small anisotropy parameters. However, in contrast to DTPA-based LCTs, the presented LCT shows only one set of signals and furthermore, the protein-tag conjugate is stable under reductive conditions and the lanthanoid complex displays a higher stability than EDTA-lanthanoid chelates.

In order to synthesize a pyridine-derived LCT that provides more rigid side-arms, Huang et al. developed the 4'-mercapto-2,2':6',2''-terpyridine-6,6''-dicarboxylic acid (4MTDA) LCT.¹³⁸ The reported LCT provides very large anisotropy parameters of $\Delta\chi_{ax} = 38.9 \times 10^{-32} \text{ m}^3$ and $\Delta\chi_{rh} = 5.2 \times 10^{-32} \text{ m}^3$ for the Dy^{3+} -loaded chelator on an ubiquitin A28C construct. The large observed PCSs and RDCs up to 13 Hz can be attributed to additional immobilization of the chelator on the surface of the protein by the carboxylate bearing sidechain of the residue E24. However, a significant drawback of the reported 4MTDA LCT is the ligation yield of only 50–70%, which is lower than for the previously described 3MDPA and 4MMDPA LCTs.

In order to provide LCTs that form a reduction-stable thioether bond upon conjugation to the protein, Yang et al. developed the (4-(phenylsulfonyl)-pyridine-2,6-dicarboxylic acid (4PS-DPA) and 4-phenylsulfonyl-(pyridine-2,6-diyl)bismethylenetrinitrilo tetrakis(acetic acid) (4PS-PyMTA)) LCTs.¹¹⁹ The reported LCTs, which are closely related to the 4VPyMTA LCT published by Yang et al.¹³⁷ but incorporate a significantly shorter and more rigid linker moiety, yield large anisotropy parameters of $\Delta\chi_{ax} = -24.5 \times 10^{-32} \text{ m}^3$ and $\Delta\chi_{rh} = -7.1 \times 10^{-32} \text{ m}^3$ for Dy^{3+} on a ubiquitin G47C R72A R74A construct.

2.6 DOTA-Based Lanthanoid Chelating Tags

Due to the excellent chelation of lanthanoids by DOTA-derived chelators (equilibrium constant for the formation of the Gd^{3+} -DOTA complex = 10^{25} M^{-1} at pH 7)¹³⁹ that constitutes the main reason for its extensive application in magnetic resonance imaging (MRI) as well as radiopharmacy,^{140,141} the macrocyclic DOTA framework presents an optimal scaffold for the introduction of various substituents, donor atoms and linker moieties. Furthermore, by three-fold installation of a negatively charged donor atom, e.g. three lactic acid side-arms, and a noncharged linker donor atom, neutral lanthanoid DOTA complexes can be conveniently synthesized and the chance for Coulomb interactions of the LCT with the protein is successfully minimized. Due to the high affinity towards lanthanoid ions in combination with the possibility to rationally design the LCTs by introduction of suitable side arms, the DOTA ligand displays very favorable properties for the synthesis of high-performance LCTs.

Since DOTA-based LCTs can, based on two conformations of the ring scaffold ($\delta\delta\delta\delta$ and $\lambda\lambda\lambda\lambda$) and two conformations of the pendant arms of the chelator (Δ and Λ), form up to four different stereoisomers in solution ($\Delta(\delta\delta\delta\delta)$, $\Delta(\lambda\lambda\lambda\lambda)$, $\Lambda(\delta\delta\delta\delta)$, $\Lambda(\lambda\lambda\lambda\lambda)$) (Fig. 10), methyl substituents were introduced in order to restrict the conformational flexibility of the ligand.

Häussinger et al. reported in 2009 an eight-fold methyl-substituted LCT (Ln-DOTA-M8-(8S)-SSPy) with a single-armed linker that leads to very large PCSs over 5 ppm and RDCs over 20 Hz, properties that were unprecedented for single-armed LCTs.²⁹ The methyl substituents on the basic ring scaffold occupy in an equatorial-upper position^{128,142} upon coordination of the metal in order to minimize steric clashes and provide a suitable cavity for the lanthanoid ion. As shown by the authors, the introduction of methyl-substituents leads to a substantial increase of the conformational stability of the ligand that can be demonstrated in variable temperature spectra. Furthermore, the LCT can be used under extreme buffer conditions as e.g. strongly acidic pH-values, thereby enabling to monitor the unfolding of a protein.²⁹ The Tm^{3+} loaded LCT reported by Häussinger et al. adopts in solution a twisted square antiprismatic $\Delta(\delta\delta\delta\delta)$ conformation with negligible amounts of the square antiprismatic $\Lambda(\delta\delta\delta\delta)$ isomer (9% based on HPLC detection), while the Dy^{3+} complex, that adopts a square antiprismatic $\Lambda(\delta\delta\delta\delta)$ geometry as major conformation, contains a more substantial amount of the minor conformational isomer (27%).¹⁴²

As reported by Joss et al., the occurrence of the minor conformational isomer can be suppressed by an inversion of the stereochemistry of the side arms that results in exclusive presence of the $\Lambda(\delta\delta\delta\delta)$ isomer in the strongly paramagnetic Tb^{3+} , Dy^{3+} and Tm^{3+} complexes as well as in the diamagnetic lutetium complex.³⁴ Subsequently, it was demonstrated by Opina et al. for the Ln-DOTA-M8-(4R4S) complexes, that the observation of the occurrence of only one conformational species, i.e. $\Lambda(\delta\delta\delta\delta)$, holds true for the whole lanthanoid series (the radioactive promethium was not investigated).¹⁴³ Surprisingly, the fitted anisotropy parameters on ubiquitin and human carbonic anhydrase II constructs both yield that the observed paramagnetic susceptibility of Ln-DOTA-M8-(4R4S)-SSPy³⁴ is significantly different from the previously published Ln-DOTA-M8-(8S)-SSPy.²⁹ Since both LCTs are structurally identical apart from the stereochemistry of the side arms and the associated switch in conformational geometry, it can be concluded that the coordination polyhedron and ligand-metal interactions can lead to a significant change in the induced anisotropy as suggested earlier for a range of coordination polyhedra by Mironov et al.¹⁰⁶

In order to expand the repertoire of DOTA-derived tags and enhance the obtained shifts with DOTA-M8-SSPy LCTs, Joss et al. introduced isopropyl substituents on the basic ring scaffold (Ln-P4M4-DOTA) and evaluated the performance of the novel chelator on ubiquitin and human carbonic anhydrase II constructs.^{36,144} The Ln-P4M4-DOTA LCT induces large PCS of up to 6.5 ppm on ubiquitin S57C and shows significantly increased anisotropy parameters when compared to its predecessor Ln-DOTA-M8-(4R4S)-SSPy. A comparison of the different properties of the newly synthesized paramagnetic complexes with Ln-DOTA-M8-(4R4S)-SSPy showed that the shift range (measured in one-dimensional ^1H spectra), position of donor atoms (monitored by DFT calculations) and translational motion of the LCT on the surface of the protein (modeled with a program developed by Suturina et al.¹⁴⁵) are highly similar for both LCTs. Therefore, it can be concluded that the reason for the strongly enhanced shifts induced by Tm- and

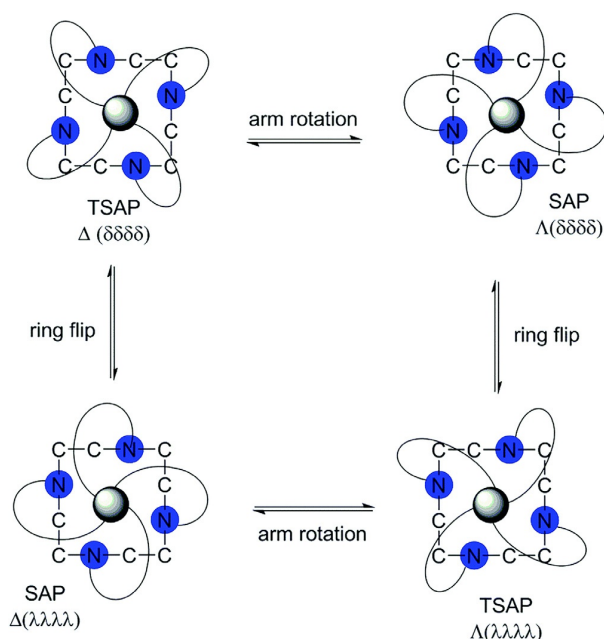


Fig. 10 Schematic representation of the different conformations of a DOTA-type ligand upon metal coordination and the paths of conversion between the conformers.¹⁴² Reproduced by permission of The Royal Society of Chemistry from Opina, A. C. L.; Strickland, M.; Lee, Y.-S.; Tjandra, N.; Byrd, R. A.; Swenson, R. E.; Vasalati, O. *Dalton Trans.* **2016**, 45, 4673–4687.

Dy-P4M4-DOTA when compared to Tm- and Dy-DOTA-M8-(4R4S)-SSPy can be attributed to decreased rotational freedom of the LCT relative to the protein due to the bulkier isopropyl substituents of the basic ring scaffold.

Since it would be highly desirable to apply rigidified DOTA chelators under reductive conditions as e.g. used for in-cell NMR, Müntener et al. developed a rigidified DOTA chelate bearing a pyridine-derived linker moiety (Ln-DOTA-(3R4S)-M7Py) that forms upon conjugation of the LCT to the protein a rigid and stable thioether bond.³³ Since the LCT showed only limited reactivity, an improved version with a fluorine substituent in *ortho* position of the leaving group was subsequently developed.³⁵

The electron-withdrawing substituent allows to stabilize the Meisenheimer-complex intermediate¹⁴⁶ which involves delocalization of negative charge within the aromatic ring during the S_NAr tagging reaction. However, the anisotropy parameters are partially averaged due to the suboptimal placement of the leaving group in the *para* position of the pyridine linker moiety. Therefore, a novel pyridinethiazole linker was developed that aligns the main axis of the induced anisotropy with the $C_{Pyr}-S_{Cys}$ bond and, thereby, leads to a favorable averaging of the induced anisotropy and hence to strongly enhanced PCSs and RDCs (Fig. 11).³⁵

In order to further enhance the obtained anisotropy parameters with the pyridinethiazole linker, the methyl substituents on the macrocyclic ring scaffold were replaced by isopropyl substituents to yield Ln-P4T-DOTA (Fig. 12).³⁷ Ln-P4T-DOTA induces very large PCS and was benchmarked on ubiquitin S57C, ubiquitin K48C and ¹⁵N selectively leucine labeled hCA II S166C constructs (anisotropy parameters: $\Delta\chi_{ax} = 44.3 \times 10^{-32} \text{ m}^3$ and $\Delta\chi_{rh} = 4.8 \times 10^{-32} \text{ m}^3$ for Tm³⁺ on a hCA II S166C construct). Due to its reduction-stable linker, the LCT could be applied for in-cell NMR studies.³⁷

In order to reduce motional freedom of the LCT on the surface of the protein, a two-point anchored LCT named CLaNP-3 was reported in 2007 by Vlasie et al.¹⁴⁷ In contrast to single-armed LCTs, two-point anchored LCTs yield in general strongly enhanced anisotropy parameters, since they are significantly more tightly attached to the surface of the protein. However, the expression of protein constructs that bear two cysteine residues in close proximity and provide thereby the attachment site of the double-armed LCT is more challenging than the expression of single cysteine mutants, since protein misfolding after expression can more easily

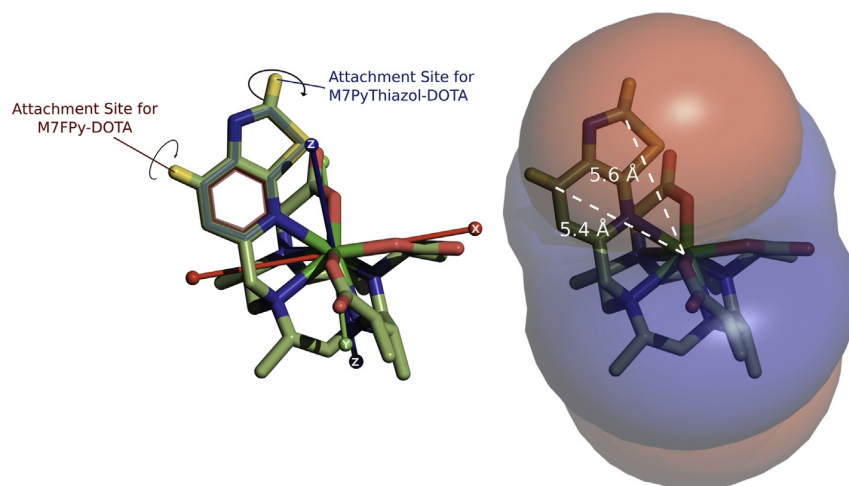


Fig. 11 Difference in the attachment point between Ln-M7FPy-DOTA and Ln-M7PyThiazol-DOTA. The more favorable alignment of the $C_{Pyr}-S_{Cys}$ linkage of Ln-M7PyThiazol-DOTA leads to a reduced averaging of the induced anisotropy and thus to larger paramagnetic effects observed on the protein.³⁵ Reproduced with permission from Müntener, T.; Kottelat, J.; Huber, A.; Häussinger, D. *Bioconjug. Chem.* **2018**, *29*, 3344–3351. Copyright 2016 American Chemical Society.

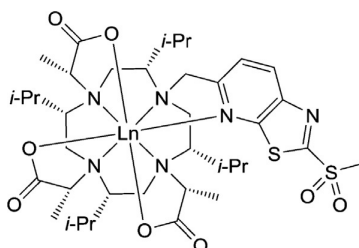


Fig. 12 Chemical structure of the $\Lambda(\delta\delta\delta\delta)$ conformer of Ln-P4T-DOTA.³⁷ The lanthanoid ion is coordinated by three carboxylate and one pyridine donors and the sterically demanding substituents rigidify the DOTA scaffold of the ligand. The methylsulfone on the pyridinethiazole moiety serves as leaving group upon reaction of the LCT with a cysteine residue of the protein.

occur than in proteins with cysteines that are located farther from each other. Using CLaNP-3 loaded with Yb^{3+} , PCSs for residues up to 35 Å away from the metal center were detected. Although the LCT presented by Vlasie et al. shows multiple conformations in solution that are in exchange with each other, in ^1H - ^{15}N HSQC experiments only one set of signals for most of the residues is detected, leading to the successful determination of one set of anisotropy parameters.

After the initial assessment of two-armed DOTA LCTs by Vlasie et al.,¹⁴⁷ Keizer et al. synthesized and reported the double-armed, DOTA-based CLaNP-5 LCT.^{28,148} The CLaNP-5 LCT presented by Keizers et al. yields large anisotropy parameters of $\Delta\chi_{\text{ax}} = 55.3 \times 10^{-32} \text{ m}^3$ and $\Delta\chi_{\text{rh}} = 6.9 \times 10^{-32} \text{ m}^3$ for Tm^{3+} and $\Delta\chi_{\text{ax}} = 9.0 \times 10^{-32} \text{ m}^3$ and $\Delta\chi_{\text{rh}} = 2.3 \times 10^{-32} \text{ m}^3$ for Yb^{3+} on a Paz E51C E54C construct.

Since the CLaNP-5 LCT displays a relatively high, three-fold positive charge, Liu et al. developed the one-fold positively charged CLaNP-7 LCT that is equipped with two nitro-substituted phenol side arms instead of the two pyridine N-oxide moieties found in CLaNP-5.¹²¹ Upon coordination of the metal, the two phenol groups are deprotonated to give the corresponding phenoxides that partially compensate the three-fold positive charge of the lanthanoid ion. Interestingly, the anisotropy parameters were found to be pH-dependent when a histidine residue is located in the immediate vicinity, a behavior that could originate in an apically coordinated water/hydroxide that can interact with the histidine sidechain.

Subsequently, Liu et al. developed the one-fold positively charged CLaNP-9 LCT that bears two carboxylate side arms.¹²² The presented LCT yields large PCS and anisotropy parameters on Paz Cu^{II} E51C E54C as well as T4 Lysozyme N55C V57C ($\Delta\chi_{\text{ax}} = 39.5 \times 10^{-32} \text{ m}^3$ and $\Delta\chi_{\text{rh}} = 10.6 \times 10^{-32} \text{ m}^3$ for Tm^{3+} on the Paz construct and $\Delta\chi_{\text{ax}} = 32.5 \times 10^{-32} \text{ m}^3$ and $\Delta\chi_{\text{rh}} = 15.6 \times 10^{-32} \text{ m}^3$ for Tm^{3+} on the T4 Lysozyme construct). Although the Yb-CLaNP-9 complex exhibits larger shifts in its one-dimensional ^1H spectrum, it displays smaller anisotropy parameters upon attachment to the protein when compared to CLaNP-5. This result suggests that the CLaNP-9 LCT shows a slightly enhanced mobility on the surface of the protein, which can be attributed to the linker that is elongated by ~ 2 Å when compared to CLaNP-5.

In order to decrease the conformational flexibility of the DOTA scaffolds, Graham et al. reported a DOTA-amide LCT that displays sterically demanding side arms that restrict the flexibility of the basic ring scaffold and side arms of the LCT and thus leads to only one set of signals in ^1H - ^{15}N HSQC experiments.³¹ It is noteworthy, that the shift range in one-dimensional ^1H spectra of the ytterbium complex amounts only to 170.5 ppm, while the LCT induces very large anisotropy parameters of $\Delta\chi_{\text{ax}} = 37.0 \times 10^{-32} \text{ m}^3$, $\Delta\chi_{\text{rh}} = 12.0 \times 10^{-32} \text{ m}^3$, for Ln = Tm^{3+} on an ArgN protein construct. This result indicates that the bulky side arms do not only lead to a decrease in flexibility of the macrocyclic scaffold, but also efficiently suppress rotational averaging of the paramagnetic effects. Besides the introduction of more steric bulk, also shortening of the linker moiety can lead to significantly enhanced paramagnetic effects observed on the protein.

Lee et al. reported compact, hydrophilic DOTA-based LCTs bearing methyl substituents on the side arms and a short linker moiety resulting in the formation of a disulfide linkage to the protein.⁶⁸ The LCT presented by Lee et al. yields large observed anisotropy parameters on ubiquitin A28C ($\Delta\chi_{\text{ax}} = -19.4 \times 10^{-32} \text{ m}^3$ and $\Delta\chi_{\text{rh}} = -7.8 \times 10^{-32} \text{ m}^3$ for Tm^{3+}) and HPPK S112C C80A ($\Delta\chi_{\text{ax}} = 54.5 \times 10^{-32} \text{ m}^3$ and $\Delta\chi_{\text{rh}} = 12.5 \times 10^{-32} \text{ m}^3$ for Tm^{3+}). In order to exploit the advantages of two-point anchored LCTs, Lee et al. converted both single-armed LCTs from the reported study,⁶⁸ i.e. the LCTs with a disulfide linker similar to Häussinger et al.²⁹ as well as the single-armed LCTs with a shorter disulfide linker, into two-point anchored LCTs that yield very large anisotropy parameters for the Tm^{3+} -loaded LCT on the tested ubiquitin E24C A28C and HPPK K76C C80 constructs (Fig. 13).^{123,124}

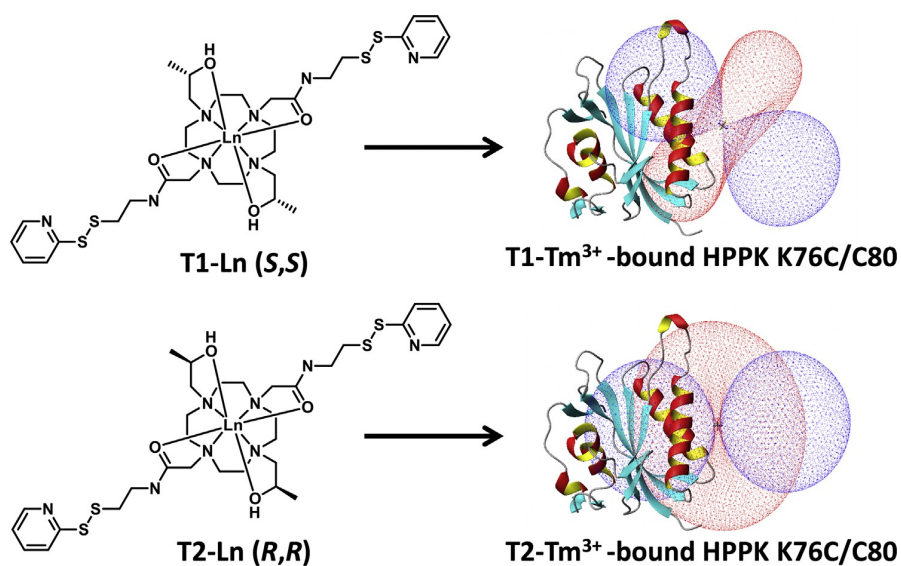


Fig. 13 Chemical structures of the Ln-T1 and Ln-T2 and isosurfaces induced by the Tm^{3+} -loaded LCTs attached to a HPPK K76C C80 construct.¹²³ Reproduced by permission of The Royal Society of Chemistry from Lee, M. D.; Dennis, M. L.; Swarbrick, J. D.; Graham, B. *Chem. Commun.* **2016**, 52, 7954–7957.

Remarkably, the magnitude of the induced PCS by the two-point anchored LCTs with a longer linker is significantly pH-dependent, i.e. the induced shifts at pH 6.5 are 40% smaller than the ones observed at pH 8.0.¹²³ This feature could be attributed to the hydroxyl donor groups, since for the related single-armed Ln-DOTA-M8-(8S)-SSPy LCTs equipped with carboxy groups, the pH dependence of the induced anisotropy was found to be not significant by Strickland et al. in the pH range of 5.0–7.4.⁷⁰ However, the pH sensitivity of the reported LCTs by Lee et al.¹²³ was not found for the LCTs with hydroxyl donor groups and a shortened linker,¹²⁴ therefore other reasons than the hydroxyl groups may be involved in the observed pH-dependence.

In order to provide reductively-stable linkages that can be applied, e.g. in in-cell NMR, the attachment of the LCT to a AzF residue incorporated in the protein via copper catalyzed azide-alkyne click reaction provides a very valuable solution.⁶⁷ Loh et al. reported two different DOTA-based LCTs with different linker lengths. For the shorter alkyne linker presented in the study, the triazole-moiety formed upon reaction of the alkyne group on the LCT with the azide side chain directly serves as donor to the lanthanoid ion after successful conjugation reaction. Both LCTs induce sizeable anisotropy parameters on a ubiquitin Glu18AzF construct with Tm³⁺ or Tb³⁺-loaded complexes. Importantly, the used protein must tolerate free copper ions used for the catalysis of the conjugation reaction. Therefore, protein constructs with His₆ tags are usually not suitable for ligation reactions with the described methodology.

A further example of an LCT yielding a reduction-stable linkage upon conjugation constitutes the iodoacetamide LCT Ln-M8-CAM-I reported by Hikone et al., which incorporates a carbamidemethyl linker with an iodide leaving group on an acetyl amide moiety that reacts as electrophile with the nucleophilic thiol group of cysteine residues of the protein.⁹⁵ In order to demonstrate the suitability of the LCT derived from Ln-DOTA-M8-(8S)-SSPy²⁹ for in-cell NMR studies, Hikone et al. ligated the newly developed LCT to ubiquitin S57C and introduced the tagged constructs successfully into HeLa cells via electroporation. The derived anisotropy parameters ($\Delta\chi_{\text{ax}} = 4.0 \times 10^{-32} \text{ m}^3$ and $\Delta\chi_{\text{rh}} = 1.8 \times 10^{-32} \text{ m}^3$ for Dy³⁺ at 310 K) demonstrate the suitability of the reported construct for in-cell NMR, however, the magnitude of the anisotropy parameters is significantly smaller than for Ln-DOTA-M8-(8S)-SSPy ($\Delta\chi_{\text{ax}} = 19 \times 10^{-32} \text{ m}^3$ and $\Delta\chi_{\text{rh}} = 11 \times 10^{-32} \text{ m}^3$ for Dy³⁺ at 313 K), which can be attributed to the significantly longer and therefore more flexible iodoacetamide linker.^{29,95}

Yang et al. presented in 2016 an LCT that yields extremely large anisotropy parameters and PCSs on an ubiquitin G47C construct ($\Delta\chi_{\text{ax}} = -84.3 \times 10^{-32} \text{ m}^3$ and $\Delta\chi_{\text{rh}} = -17.2 \times 10^{-32} \text{ m}^3$ for Dy³⁺ and $\Delta\chi_{\text{ax}} = 65.2 \times 10^{-32} \text{ m}^3$ and $\Delta\chi_{\text{rh}} = 31.8 \times 10^{-32} \text{ m}^3$ for Tm³⁺).³² The large PCSs are highly desirable for evaluation of challenging applications involving large biomacromolecules, however, due to the nonstabilized Meisenheimer complex¹⁴⁶ during tagging reaction of the cysteine with the pyridinesulfone moiety, the LCT requires severely basic tagging conditions, i.e. pH 9.0–9.3 over a time of 16–18 h. For a successful application of the strongly paramagnetic LCT, only proteins withstanding these harsh conditions can be investigated.

In order to attach LCTs to other biomacromolecules than proteins, Wu et al. developed a bromoacetamide linker that was demonstrated to be able to ligate to DNA incorporating a phosphorothioate modification.⁷² However, conjugation to an internal phosphorothioate in DNA generates an additional stereocenter, i.e. a stereogenic phosphorus atom, in racemic form and hence to two sets of diastereotopic NMR signals. In the reported case, the two diastereomers can be successfully separated using conventional reverse-phase HPLC. PCSs were extracted from 2D NOESY spectra and fitted to a structural model of the DNA double-helix based on fiber diffraction data in order to give the first anisotropy parameters that were obtained by using an LCT with a covalent linkage to DNA.

The discussed examples in this section highlight the outstanding properties of DOTA-type ligands for the development of tuneable, strongly paramagnetic LCTs with a diverse range of linker and activator moieties and their subsequent application in structural studies of protein–ligand complexes.

3 Applications of Lanthanoid Chelating Tags in the Localization of Ligands Within Biomacromolecules

3.1 Introduction

The localization of ligands within biomacromolecules is crucial for the understanding of the cellular machinery and the underlying processes. Furthermore, the positioning and orientation of a ligand within a drug target is an important prerequisite for rational design of drug molecules. In contrast to X-ray crystallography, which provides in principle a precise tool for the determination of ligand-protein complexes, the use of NMR spectroscopy allows for monitoring of biomacromolecule-ligand interactions under physiologically relevant conditions. While NOESY experiments are optimally suited for monitoring distances below 5–8 Å and require a full NMR assignment obtained from multidimensional experiments by using isotopically labeled biopolymers, CSP can result in misleading conclusions due to conformational changes or unexpected intermolecular interactions of the biomacromolecules. The exploitation of PCS for the characterization of biomacromolecule-ligand complexes can cover large distances beyond 100 Å with high accuracy. Furthermore, if the anisotropy parameters from multiple sites have been determined, the accordingly tagged protein constructs can be used with a large number of ligands, thus providing a valuable tool for screening of potential drug molecules.

3.2 Weakly Binding Ligands

In order to determine the position of a weakly binding ligand within a protein that is in fast exchange, the ligand shifts can be conveniently obtained from one-dimensional NMR experiments. Since the ligand can be employed in excess, the signal discrimination between ligand resonances and protein peaks is ensured by the intensity and the PCS can be extracted from the observed

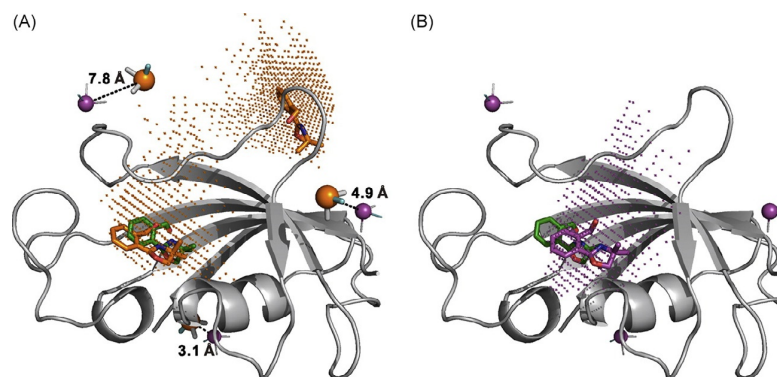


Fig. 14 Localization of ligands within a protein scaffold (predicted metal positions: magenta, experimentally determined metal positions: orange, NOE-based ligand positions: green, ligand positions determined by PCS: orange, ligand position determined by use of predicted anisotropy parameters: magenta). Fitted ligand position based on experimental data is in accordance to NOE position, with simultaneous occurrence of a “ghost” site during the fitting process (A). Fitted ligand position based on predicted anisotropy parameters (B).⁴⁰ Reproduced with permission from Guan, J.-Y.; Keizers, P. H. J.; Liu, W.-M.; Löhr, F.; Skinner, S. P.; Heeneman, E. A.; Schwalbe, H.; Ubink, M.; Siegal, G. *J. Am. Chem. Soc.* **2013**, *135*, 5859–5868. Copyright 2013 American Chemical Society.

shift for the fast exchanging bound and unbound fraction of the ligand provided the affinity is known, and the PCS is large compared to the CSP. Saio et al. determined the positioning of low affinity ligands within the growth factor receptor-bound protein 2 Src homology domain by use of PCS analysis.¹⁰² The paramagnetic shifts were induced by a lanthanoid-binding peptide tag attached via two linkages on the N-terminus and one cysteine mutant in the protein. Upon determination of the anisotropy parameters, the weakly binding pYTN tripeptide ligand and the tagged protein construct were combined and the PCS were extracted from the determined averaged shift of the unbound and the bound state using the binding affinity of the tripeptide to the protein obtained from a ^1H - ^{15}N HSQC-based titration.

The combination of the structural restraints obtained by PCS analysis and docking simulations yielded then the position of the ligand within the biomacromolecule. In 2013, Guan et al. presented the localization of ligands using PCS analysis using the rigidly, two-point anchored LCT CLaNP-5 attached to three different sites of the protein.⁴⁰ Interestingly, two-point anchored LCTs as CLaNP-5 that are strongly immobilized on the surface of the protein, yield reproducible and predictable anisotropy parameters (Fig. 14). This leads to the convenient feature that the predicted $\Delta\chi$ parameters can be directly used for a coarse grain localization of the ligand within the biomacromolecule (RMSD of 4.4–4.7 Å relative to NOE structure). Hence, for an initial assessment or discrimination of two already established binding sites in a protein, no isotopically labeled protein or backbone amide assignment is required and the only input required is the PCS from one-dimensional ^1H experiments. The authors reported that upon the paramagnetic labeling of the protein on three different sites and fitting of the anisotropy parameters based on backbone amide NMR shifts, the ligand can be localized and positioned over a distance of 15–25 Å with an average root-mean-square deviation (RMSD) of 2.8 Å when compared to the NOE structure. In contrast to the attachment of the paramagnetic tag to the biomacromolecule of interest, as demonstrated by Brath et al., the LCT can also be directly attached to the weakly binding ligand.¹⁴⁹ Thereby, the authors identified the interaction site of calmodulin with a low-affinity anesthetic agent. However, although the direct attachment of the LCT to the ligand of interest has the advantage that dissolution of hydrophobic ligands is enhanced in aqueous solution and engineering of protein cysteine mutants is omitted, the approach reported by Brath et al. is less suitable for the screening of a large number of different ligands, since the LCT would have to be synthetically coupled to each of the applied ligands.

3.3 Ligands in the Intermediate Exchange Regime

Ligands in the intermediate exchange regime can pose significant challenges for PCS NMR approaches, since their resonances are often broadened beyond detection and therefore no structural restraints can be measured. An elegant tool to circumvent this difficulty is the use of the chemical saturation exchange transfer methodology (CEST) as proposed by Gao et al.¹⁵⁰ The authors of the study suggested to detect the PCS on the spectrum of the free ligand by scanning through the ^{19}F saturation frequencies of the bound state. Since intensity dips can be observed at the frequencies of the signals of the bound ligand, a fluorine spectrum of the bound ligand can thereby be collected. The authors of the study then used the PCS recorded by use of their ^{19}F CEST methodology in combination with PCS-driven docking programs as HADDOCK or XPLOR-NIH containing the PARArestraints module in order to localize mono- and difluorinated inhibitors in the BRM bromodomain.

In order to extend the presented method by Gao et al.¹⁵⁰ for the localization of ligands with micromolar affinities and NMR resonances in the intermediate exchange regime that do not incorporate fluorine, Xu et al. developed a new method that is suitable for localization of ligands with micromolar affinities.¹⁵¹ Since ^1H CEST usually is not applicable in a clean fashion, since it can suffer severely from NOE interferences, Xu et al. proposed the use of relaxation dispersion (Fig. 15). By application of their novel methodology, the authors obtained PCS restraints and used them successfully as input restraints in docking simulations of the ligand within the protein.

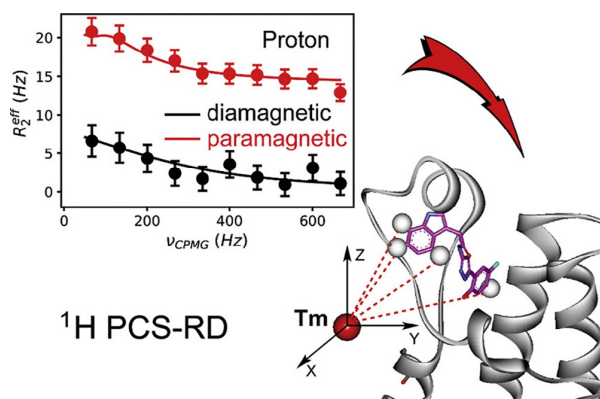


Fig. 15 Extraction of PCS measured on ligands in the intermediate exchange regime by use of relaxation dispersion.¹⁵¹ Reproduced with permission from Xu, D.; Li, B.; Gao, J.; Liu, Z.; Niu, X.; Nshogoza, G.; Zhang, J.; Wu, J.; Su, X.-C.; He, W.; Ma, R.; Yang, D.; Ruan, K. J. *Phys. Chem. Lett.* **2018**, *9*, 3361–3367. Copyright 2018 American Chemical Society.

3.4 Strongly Binding Ligands

In the localization of strong ligands, usually large PCS can be detected, since the observed signal originates exclusively from the bound state that experiences the full PCS and not an average shift of the bound and unbound state as in the case of the weakly binding ligands in fast exchange. However, the signal discrimination of the bound, strongly binding ligand against the signals of the protein, which is present in equimolar amounts within the sample, bears a significant challenge. Besides the equimolar ratio and therefore similar shift intensities, tightly binding ligands exhibit the same rotational correlation time as the signals of the protein and are therefore broadened in similar extent. Isotopic labeling of the ligand, e.g. with ^{13}C atoms, is certainly achievable, however, for a large number of ligands that have to be screened against a protein of interest this cannot be considered as a convenient approach. Saio et al. achieved the required signal discrimination by deuteration of the protein and measurement of one-dimensional ^1H spectra.¹⁰² Thereby, the authors were able to localize a high-affinity inhibitor of the Src homology 2 domain of the growth factor receptor-bound protein 2 by use of the observed PCS and docking simulations over a distance of ~ 10 Å.

An approach for overcoming the issue of signal discrimination of strongly binding ligands was proposed by Chen et al., who used a *tert*-butyl group, which displays a strong signal with a specific NMR shift, as reporter group for detection of the signal of the bound ligand (Fig. 16).¹⁵²

The researchers found that broadening and simultaneous shift of the signals due to the paramagnetic tags can lead to ambiguous assignments caused by insufficient signal discrimination when compared to the protein. However, the *tert*-butyl group incorporated in the ligand produces very intense NOESY correlations even when the protein is hardly detectable. Thereby, the signals of the ligand can be unambiguously assigned and discriminated towards the protein background. Since the used approach omits the need for perdeuteration of the protein samples and spectra can be acquired at very low concentration, which is important for screening of a large number of ligands, the methodology reported by Chen et al. is highly suitable for all ligands containing *tert*-butyl groups. In order to unambiguously localize fluorine-containing ligands over large distances, Zimmermann et al. developed an approach based on the acquisition of one-dimensional ^{19}F spectra with paramagnetic tags attached to different mutation sites on the protein scaffold.⁴¹ This methodology will be described in the next section, which covers the enzyme human carbonic anhydrase II and the localization of ligands within the protein by PCS, and lies at the heart of the present case study.

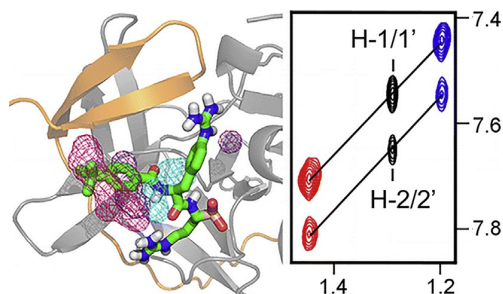


Fig. 16 Strong correlation peaks of the *tert*-butyl group of the applied, strongly binding ligand (right). Structural models of the protein–ligand complex of DENpro and ligand 1. The fitted regions with a deviation less than 0.02 ppm from the experimental PCSs are indicated with color coding: *tert*-butyl group (magenta), aromatic protons (purple and cyan).¹⁵² Reproduced with permission from Chen, W.-N.; Nitsche, C.; Pilla, K. B.; Graham, B.; Huber, T.; Klein, C. D.; Otting, G. *J. Am. Chem. Soc.* **2016**, *138*, 4539–4546. Copyright 2016 American Chemical Society.

4 Localization of Ligands Within Human Carbonic Anhydrase II Using ^{19}F Pseudocontact Shift Analysis

4.1 Human Carbonic Anhydrase and Its Use in the Development of Artificial Metalloenzymes

Human carbonic anhydrase II (hCA II) is a medium-sized, monomeric, globular protein with a molecular weight of 29 kDa. hCA II catalyzes the reversible hydration of CO_2 to bicarbonate and displays a catalytic turnover rate of $k_{\text{cat}} = 1.4 \times 10^6 \text{ s}^{-1}$ and an efficiency of $k_{\text{cat}}/K_{\text{M}}$ of $1.5 \times 10^8 \text{ M}^{-1} \text{ s}^{-1}$, which rank among the highest known for enzymatic reactions and can be considered to be close to the limit imposed to the reaction by diffusion.^{153–156} The reaction is best described by a two-step mechanism.¹⁵⁶ First, CO_2 is directly attacked by the zinc bound hydroxide ion to form a metal bound bicarbonate, which is then replaced by a water molecule in order to release the bicarbonate anion from the zinc metal center. In the second step, a proton from the zinc bound water molecule is transferred, using a histidine residue as proton shuttle, to the aqueous solvent surroundings in order to regenerate the zinc bound hydroxide. The central zinc(II) ion of hCA II is coordinated by three histidine residues and one hydroxide ion in a distorted tetrahedral geometry in order to form together the active site. The active site itself is located at the bottom of a conical shaped cavity, which is about 15 Å deep and 15 Å wide at the entrance. The cavity leading to the active site comprises a hydrophilic and a hydrophobic side.¹⁵⁷ The Zn(II) ion in combination with the hydrophobic side is the binding site for virtually all hCA II inhibitors.¹⁵⁶ An interesting structural feature displayed by hCA II is the knot-like topology, i.e. if one were to pull at both ends of the peptide chain, a knot would form (Fig. 17).¹⁵⁷

In order to use hCA II as a drug target, high affinity inhibitors of hCA II have been developed.¹⁵⁶ Strongly binding inhibitors of hCA II are based on an aromatic sulfonamide functional group within the ligand.^{158,159} The strong binding affinity originates in the ability of the sulfonamide anion ($\text{RS}(\text{O})_2\text{NH}^-$) to strongly coordinate the zinc ion.¹⁶⁰ The properties of the inhibitor can be tailored by adjusting the aromatic ring and adding different functional groups to the aromatic scaffold.¹⁵⁶ Importantly, the inhibition of hCA II allows the control of different diseases, e.g. glaucoma, a disease that results in nerve damage followed by a loss of the patient's vision. hCA II plays an important role in the control of the intraocular pressure. The conversion of CO_2 into bicarbonate catalyzed by hCA II in the ciliary body causes the uptake of sodium ions and thus also water into the eye. The sodium ions together with water, other electrolytes and components as amino acids and glutathione form the aqueous humor. Increased intraocular pressure is one of the most important risk factors for glaucoma. The inhibition of hCA II suppresses the formation of bicarbonate, which results in less uptake of sodium ions and water and thus leads to a reduction of the intraocular pressure.¹⁶¹

The large binding pocket of hCA II in combination with strongly binding and readily derivatizable inhibitors rendered hCA II likely to be an optimal scaffold for artificial metalloenzymes. Additionally, hCA II is easily overexpressed and is therefore highly suitable for an optimization by directed evolution in order to improve the performance of the organometallic protein construct. By today, directed evolution has become a highly advanced technique that can be readily used in order to optimize the catalytic properties of computationally designed enzyme folds or of proteins with promiscuous activity that are auspicious catalysts for other than only their natively catalyzed reaction.^{162–164} Transition metal complexes can be anchored to hCA II by tethering a bidentate ligand to an aryl-sulfonamide inhibitor.¹⁶⁵ Iridium-based piano stool complexes incorporated into hCA II were then shown to be able to catalyze an asymmetric transfer hydrogenation of imines.¹⁶⁶ Subsequently, the catalytic performance was improved by computational design, which led both to an increase of the enantioselectivity (70–96% ee) as well as the turnover number (9–59) (Fig. 18).¹⁶⁷

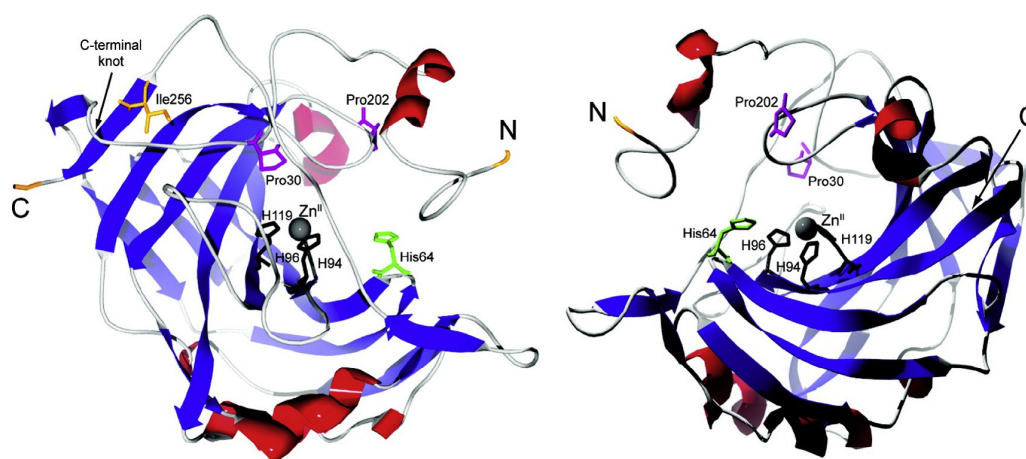


Fig. 17 View on the fold of hCA II displayed as ribbon structure (α -helices: red, β -sheets: blue). The active site Zn(II) ion is coordinated by three histidine residues, and lies at the center of the enzyme. Furthermore, the critical residues for the initiation of folding of hCA II are marked within the displayed structure.¹⁵⁶ Reproduced with permission from Krishnamurthy, V. M.; Kaufman, G. K.; Urbach, A. R.; Gitlin, I.; Gudiksen, K. L.; Weibel, D. B.; Whitesides, G. M. *Chem. Rev.* **2008**, *108*, 946–1051. Copyright 2016 American Chemical Society.

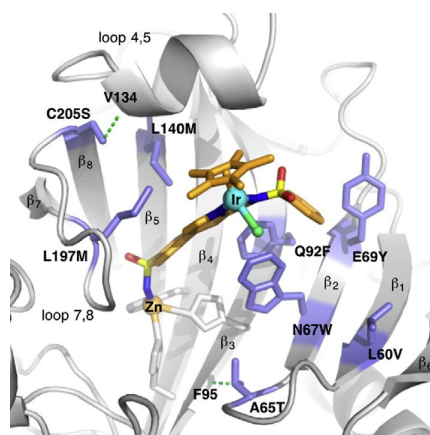


Fig. 18 View on the active site of an artificial metalloenzyme incorporating an sulfonamide iridium piano stool complex for efficient, asymmetric catalysis in the synthesis of (*S*)-salsolidine, an inhibitor of monoamine oxidase. The H-bonds introduced using computational design are marked with dashed lines.¹⁶⁷ Reproduced with permission from Heinisch, T.; Pellizzoni, M.; Dürrenberger, M.; Tinberg, C. E.; Köhler, V.; Klehr, J.; Häussinger, D.; Baker, D.; Ward, T. R. *J. Am. Chem. Soc.* **2015**, *137*, 10414–10419. Copyright 2016 American Chemical Society.

In a following study, *in vivo* transfer hydrogenation enabled by iridium catalysts incorporated into the hCA II scaffold was investigated.¹⁶⁸ Furthermore, computational studies elucidating the amino acid residues, which contribute significantly to the binding of the ligands,¹⁵⁹ as well as comparing both point-charge and atomic multipole-scored methods for studying the binding of sulfonamide ligands to hCA II were reported.¹⁶⁹ One of the most promising fields of applications for metalloenzymes is their ability to activate drugs at a specific target location *in vivo*. This could be used to specifically target cancer cells, since cancer cells overexpress carbonic anhydrase IX in order to arrange with acidosis due to their strongly accelerated and anaerobic metabolism.¹⁷⁰ In this context, the ring-closing metathesis (RCM) constitutes a promising reaction as it allows the bioorthogonal conversion of acyclic diolefins to cyclic products. By use of hCA-based artificial metalloenzymes, RCM can be performed by anchoring a second-generation Hoveyda-Grubbs catalyst to the active site of the protein using an arylsulfonamide part on the catalyst.¹⁷¹

The same properties that render hCA II as an interesting host for metalloenzymes also result in an excellent suitability as test protein for biophysical studies of protein–ligand binding.¹⁵⁶ The convenient derivatization of the inhibitors in combination with reported X-ray structures of protein–ligand complexes pose hCA II as an excellent choice in order to develop new methods of ligand localization. In order to localize ligands for hCA II within the protein scaffold in solution, a method based on paramagnetic NMR spectroscopy was developed.⁴¹ More specifically, ¹⁹F detected, lanthanoid-induced PCSs served as structural restraints in order to localize the ligands of interest within the protein scaffold in solution resulting in an accuracy of up to 0.8 Å over a distance of 22–38 Å.

4.2 Localization of Ligands within Human Carbonic Anhydrase II—Introduction and Concept

Localization of ligands within proteins in solution is crucial to the rational design of drugs and investigation of biomedical pathways within the cellular machinery. In order to study the structure of ligand–protein complexes that are not straightforward amenable to X-ray crystallography¹⁷² as well as cryo-electron microscopy¹⁷³ and in order to get insights into the dynamics of the complexes, NMR spectroscopy and in particular paramagnetic NMR spectroscopy constitutes a highly promising method.¹⁸ As discussed in the previous section, the localization of firmly binding ligands within proteins in solution using NMR poses significant challenges as signal discrimination and the short distance range covered by conventional NMR spectroscopy. Given the three-dimensional structural restraints and distance dependency of R^{-3} displayed by PCSs, they deliver an optimal tool for the calculation of the ligand position within the protein based on long-range restraints and can thereby overcome the distance limitation of NOE experiments and unspecific results by CSP. Furthermore, during the screening process of the ligands, the protein is not required to be labeled, since only the ligand PCS is used, detected from one-dimensional spectra. In order to achieve signal discrimination between the signals of a strongly binding ligand and the protein, different approaches were already implemented. Saio et al. used deuterated protein samples in order to conveniently detect the signals of the ligand.¹⁰² In order to remove the need for protein labeling, Chen et al. proposed to use a *tert*-butyl substituent as suitable reporter group, since in NOESY spectra exceedingly strong correlations are detected for the *tert*-butyl group, even when the protein correlations are barely visible.¹⁵² Since the percentage of drugs incorporating fluorine amounted to 25% in 2014¹⁷⁴ and the fluorine detection removes the need for protein labeling while giving an astonishingly high sensitivity due to the favorable gyromagnetic ratio and the 100% natural abundance of the NMR active isotope, we envisioned in our study to introduce a fluorine-detected localization of ligands within proteins using PCSs as structural restraints.⁴¹ In order to test our approach, we incorporated three different, strongly binding sulfonamide ligands within hCA II (Fig. 19).

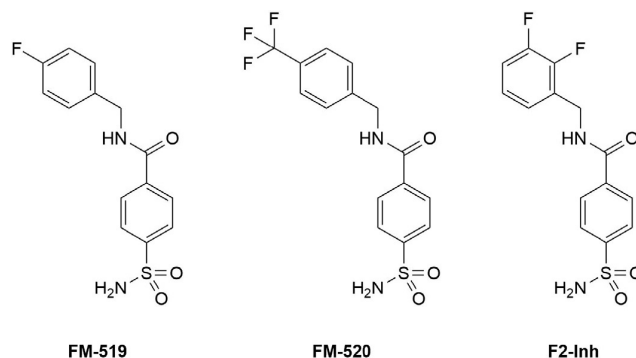


Fig. 19 hCA II inhibitors for the evaluation of the ^{19}F -PCS-based approach for localization of ligands within proteins.

For the unambiguous localization of a ligand within a protein scaffold in three-dimensional space, PCSs from different tagging sites are needed. Alternatively, a smaller number of tagging sites can be used in combination with different LCTs that yield independent anisotropy tensors. By using multiple tagging sites, first, the anisotropy parameters for each tagging site are fitted by using the backbone amides. Upon detection of the PCSs on the ligand signals, the values are compared with the back-calculated positions for a given PCS derived from the fitted anisotropy parameters depicted as isosurfaces around the paramagnetic center.

Intersection of two isosurfaces from two different positions yields a curve that includes the possible ligand positions (Fig. 20). By use of a third isosurface, the curve gets intersected at two different points that depict the remaining two possible positions for the ligand position. Intersection with a fourth isosurface leads then to one point in space as unique solution for the fluorine atom. However, since monitoring the ligand position from three different tagging sites results in two possible locations and often one of these possibilities is located completely outside the protein scaffold or in very dense regions in the interior of the protein, in practice usually a fourth position is not required to determine unambiguously the position of the ligand within the protein.

In order to obtain different tagging sites for attachment of the LCTs, five serine to cysteine mutations were introduced on the surface of the protein: S50C, S166C, S173C, S217C and S220C. Furthermore, different labeling schemes were introduced, each with its own purpose: uniformly ^{15}N labeled hCA II for the determination of the anisotropy parameters from ^1H - ^{15}N HSQC spectra and selectively ^{15}N leucine labeled hCA II in order to serve as starting point for the fitting process due to crowded paramagnetic spectra. The 26 leucine residues in hCA II show a favorable distribution in the primary sequence, amount to 10% abundance and the chemical shift dispersion in ^1H - ^{15}N HSQC experiments is reasonable as expected for a well-folded, globular protein. Upon tagging with a diamagnetic LCT, a small number of residues close to the tagging site are shifted. In order to contribute for the occurrence of these shifts, we attached, besides the paramagnetic Tm^{3+} -loaded LCT, also the diamagnetic Lu^{3+} -loaded LCT to the protein in order to acquire reference spectra. Furthermore, in order to provide complete evidence for the assignment of the shifted residues

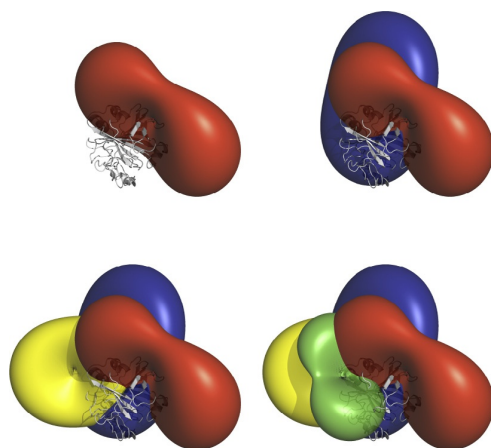


Fig. 20 Intersection of different isosurfaces for the localization of a specific nucleus within a protein scaffold using experimental and back-calculated PCS: A single isosurface plotted besides hCA II (*top left*), intersection of two isosurfaces originating from PCS measured from two different sites within the protein and their curve of intersection (*top right*), intersection of three isosurfaces resulting in two intersection points (*bottom left*, first intersection point observable in the center of the protein scaffold, the second on the backside of the protein), intersection of four isosurfaces resulting in one and thus unambiguous intersection point (*bottom right*).⁴¹ Reproduced by permission of The Royal Society of Chemistry from Zimmermann, K.; Joss, D.; Müntener, T.; Nogueira, E. S.; Schäfer, M.; Knörr, L.; Monnard, F. W.; Häussinger, D. *Chem. Sci.* **2019**, *10*, 5064–5072.

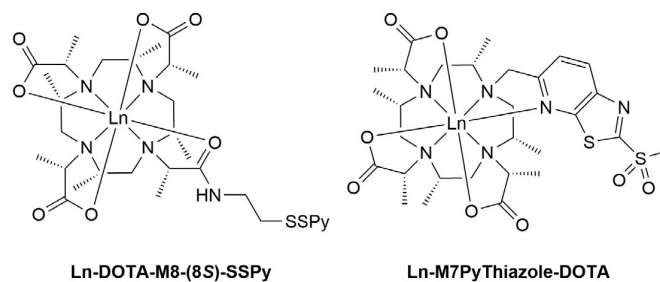


Fig. 21 Strongly paramagnetic LCTs applied in the localization of ligands within hCA II. Ln-DOTA-M8-(8S)-SSPy in $\Delta(\delta\delta\delta\delta)$ and Ln-M7-(3*R*,4*S*)-PyThiazole-DOTA in $\Lambda(\delta\delta\delta\delta)$ conformation (see also Section 2.6).^{29,35}

upon tagging reaction, we produced a triple labeled mutant that was then used for backbone assignment with and without attached Lu^{3+} -LCT.

Upon a successful assignment of the backbone of hCA II, the anisotropy parameters induced by the LCTs ligated to different attachment sites can be determined by fitting the amide signals detected in ^1H - ^{15}N HSQC experiments to the X-ray structure of hCA II. Since the uniformly ^{15}N labeled protein displays crowded spectra, selectively ^{15}N leucine labeled mutants are used to provide a reasonable starting point for the fitting routine. Full assignment of the PCS and fitting to the X-ray structure leads then to the derived anisotropy parameters. After fitting of the anisotropy parameters, the protein is incubated with the ligand of interest and excess ligand is subsequently removed by simple centrifugation. By acquiring one-dimensional ^{19}F experiments, the PCS of the fluorine atoms incorporated in the ligand can be measured. Back-calculation of the expected values in different positions in the protein by use of the previously fitted anisotropy parameters from multiple tagging sites leads then straightforwardly to the positioning of the ligand within the protein. Since large PCS decrease significantly the error and lead to more precise structural restraints, an LCT is required that induces large PCS and is suitable for exploration of the hCA II scaffold that is characterized by a diameter of 40–56 Å and a molecular weight of 30 kDa.¹⁵⁶ We used the eight-fold methyl-substituted Ln-DOTA-M8-(8S)-SSPy²⁹ and the recently published Ln-M7PyThiazole-DOTA,³⁵ which incorporates a short and rigid pyridinethiazole-derived linker moiety (Fig. 21).

4.3 PCS Analysis and Determination of the Anisotropy Parameters from Multiple Positions Within the Protein

The PCS and the corresponding anisotropy parameters are generated by the attachment of an LCT to the protein scaffold on multiple positions. In order to generate valuable structural restraints, a few criteria have to be met: (i) The LCT needs to be attached in a stable and rigid position relative to the protein. (ii) The global fold of the protein must not be compromised by the attachment of the LCT. (iii) The attachment sites should be distributed all over the protein in order to achieve a reasonable orthogonality of the different induced anisotropy parameters and the corresponding isosurfaces. In order to meet all of the mentioned criteria, five solvent-exposed serine residues in various positions of the protein located in stable secondary structure elements were mutated in order to produce five different cysteine mutants (Fig. 22). The native cysteine 206 present in human carbonic anhydrase II was mutated vice versa into a serine residue.

The PCS methodology requires a backbone assignment of the protein, in the best case a backbone assignment of the diamagnetically tagged protein in order to avoid ambiguities caused by shifted residues close to the tagging site, and therefore a backbone assignment of uniformly ^2H ^{13}C ^{15}N labeled hCA II S50C Lu-DOTA-M8 was performed. With the assigned amide resonances for the diamagnetic spectra in hand, the shifted residues due to PCSs are assigned. Based on the experimentally observed PCS on the amide resonances from multiple tagging sites and the structure of the protein, the anisotropy parameters induced by the LCT for each tagging site can then be fitted. However, due to the crowded ^1H - ^{15}N HSQC spectra of uniformly ^{15}N labeled hCA II, the direct assignment of all PCS can be cumbersome (Fig. 23).

In order to reduce the chance of obtaining an erroneous PCS assignment, multiple approaches were developed: (i) selectively labeled proteins, (ii) variable temperature NMR or (iii) use of different lanthanoids.

Selectively labeled proteins, e.g. selectively ^{15}N leucine labeled proteins as used in this study, decrease drastically the complexity of the ^1H - ^{15}N HSQC spectra (Fig. 24). The assignment of the shifted resonances in the spectra with reduced complexity that are therefore more amenable to PCS analysis leads then to an initial set of anisotropy parameters. The back-calculation of the remaining unassigned PCS based on the initially derived anisotropy parameters then strongly facilitates the complete assignment.

Measurement of PCS at variable temperatures produces a series of differently shifted signals, since PCSs are temperature dependent and, in addition, an increase in temperature leads to enhanced movements of the LCT on the surface of the protein and flexibility of the protein scaffold, which results in decreased PCS. The thereby produced pattern of signals significantly simplifies the assignment due to the “linkages” provided between the diamagnetic shift and the paramagnetic shift at room temperature (Fig. 25).

A very similar approach involves the application of different lanthanoid ions. The major determinants of the strength of the induced PCS are the intrinsic anisotropy of a lanthanoid within a given ligand-field/chelator and its motion on the surface of the protein. Since the motion can be expected to be similar for all different lanthanoids, except for very rare cases (when a larger

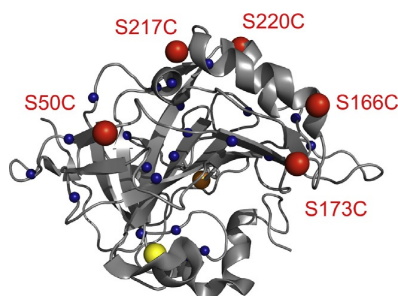


Fig. 22 Cysteine mutations introduced in the hCA II samples visualized in the X-ray structure of hCA II (PDB 3KS3,¹⁷⁵ introduced cysteine mutations: red, native cysteine 206 (mutated to serine): yellow, leucine residues: blue, zinc(II) ion: brown).⁴¹ Reproduced by permission of The Royal Society of Chemistry from Zimmermann, K.; Joss, D.; Müntener, T.; Nogueira, E. S.; Schäfer, M.; Knörr, L.; Monnard, F. W.; Häussinger, D. *Chem. Sci.* **2019**, *10*, 5064–5072.

lanthanoid ion would enable coordination of a protein sidechain on the ninth coordination site and a smaller lanthanoid would not show this effect), the complexation of different lanthanoids and attachment of the different LCTs to the protein leads to a pattern of signals similar to the measurement at various temperatures. The pattern of the shifts induced by application of different lanthanoid ions can then be used for a convenient assignment of the PCSs (Fig. 26).

The selective labeling approach is usually for severely crowded ¹H-¹⁵N HSQC spectra the more convenient choice, since the variable temperature approach and application of a series of lanthanoids even increase the complexity of the overlaid spectra in strict technical terms.

Therefore, selectively ¹⁵N leucine labeled hCA II mutants were expressed and led in combination with the assignment of uniformly labeled hCA II tagged with Lu-DOTA-M8 to an initial fitting of the acquired PCS. Based on the leucine derived tensor parameters, the PCS of the remaining residues of the uniformly ¹⁵N labeled hCA II were back-calculated and the experimentally measured PCS fitted to the structure (PDB 3KS3)¹⁷⁵ using the software Numbat.⁹⁷ A full description of the fitting routines can be found in the supporting information of the original publication.⁴¹

The complete results of all hCA II constructs tagged with Tm-DOTA-M8 are depicted in Table 1. Importantly, in order to obtain an estimation of the error included in the fitted anisotropy parameters, two different Monte-Carlo approaches can be applied. While the first one uses a structure variation during the fitting process, the second one uses only subsets of PCS to account for the possibility that erroneous assignments are hidden due to the good quality of the rest of the fitted resonances. Subsequently, two different criteria can be used to assess the quality of the fits. First, the location of the metal center obtained in a fit without used restraints in a reasonable positioning to the cysteine residue and in a distance of 6–8 Å to the γ-oxygen confirms the metal position found by the fitting routine. Second, the Q-factors give an impression about the quality of an obtained fit. The Q-factors for the used LCTs^{29,35} and hCA II constructs were found to be excellent (0.035–0.106). The complete fits of all PCS observed with the Tm-DOTA-M8-SSPy²⁹ and Tm-Thiazole³⁵ LCT constitutes the basis for an accurate back-calculation/prediction of the ¹⁹F PCS observed on the ligands and thus their localization within the protein. For the Tm tagged S166C mutant, significantly larger PCS and anisotropy parameters were found than for the other constructs. This result can be assigned to a different environment of the tagging site that causes a different flexibility of the LCT on the protein.¹⁴⁵

With the full set of fitted PCS from various attachment sites in hand, the localization of the ligands can be tested by incubation of the tagged hCA II constructs with the three different ligands. The high-affinity phenyl-sulfonamide ligand molecules (FM-519, FM-510 and F2-Inh) were added to the buffered protein sample as a solution in DMSO in small excess (1.1 eq.) and the incubated protein was then washed several times with buffer in order to remove any excess ligand. The protein was then concentrated and subjected to ¹H-¹⁵N HSQC experiments that confirmed the complete loading and the structural integrity of the protein.

4.4 Measurement of ¹⁹F PCS and Localization of the Ligand Using a Monte-Carlo Approach

¹⁹F PCS were observed in simple one-dimensional ¹⁹F experiments for all ligand-containing protein constructs (Fig. 27). Due to the strongly paramagnetic LCTs, sizeable PCS of up to 0.4 ppm were detected, ensuring that only a minimal error is included in the PCS that serves as raw data input for the fitting process of the ligand position. The linewidths of the fluorine signals in the range of 20–70 Hz indicate clearly residual mobility of the phenyl ring of the ligand that points out of the enzyme's pocket, since comparable strongly binding carazolol ligands showed linewidths of 220–230 Hz in a study of Eddy et al.¹⁷⁶ and the Debye-Waller factors of the PDB 1G54 agree with this observation.¹⁷⁷ As for the amide N—H spin pairs, the largest PCSs were observed for the Tm-tagged hCA II S166C constructs, most likely due to the suppressed flexibility of the LCT attached to this mutant when compared to the other attachment sites.¹⁴⁵

By using the obtained anisotropy parameters from the fitting to the backbone amides of the hCA II constructs, the expected ¹⁹F PCS can be back-calculated and an estimate of the position of the ligand can be achieved. The expected PCS for given anisotropy parameters in relation to the distance and angular position of the fluorine atom of the ligand can be calculated using the following equation:

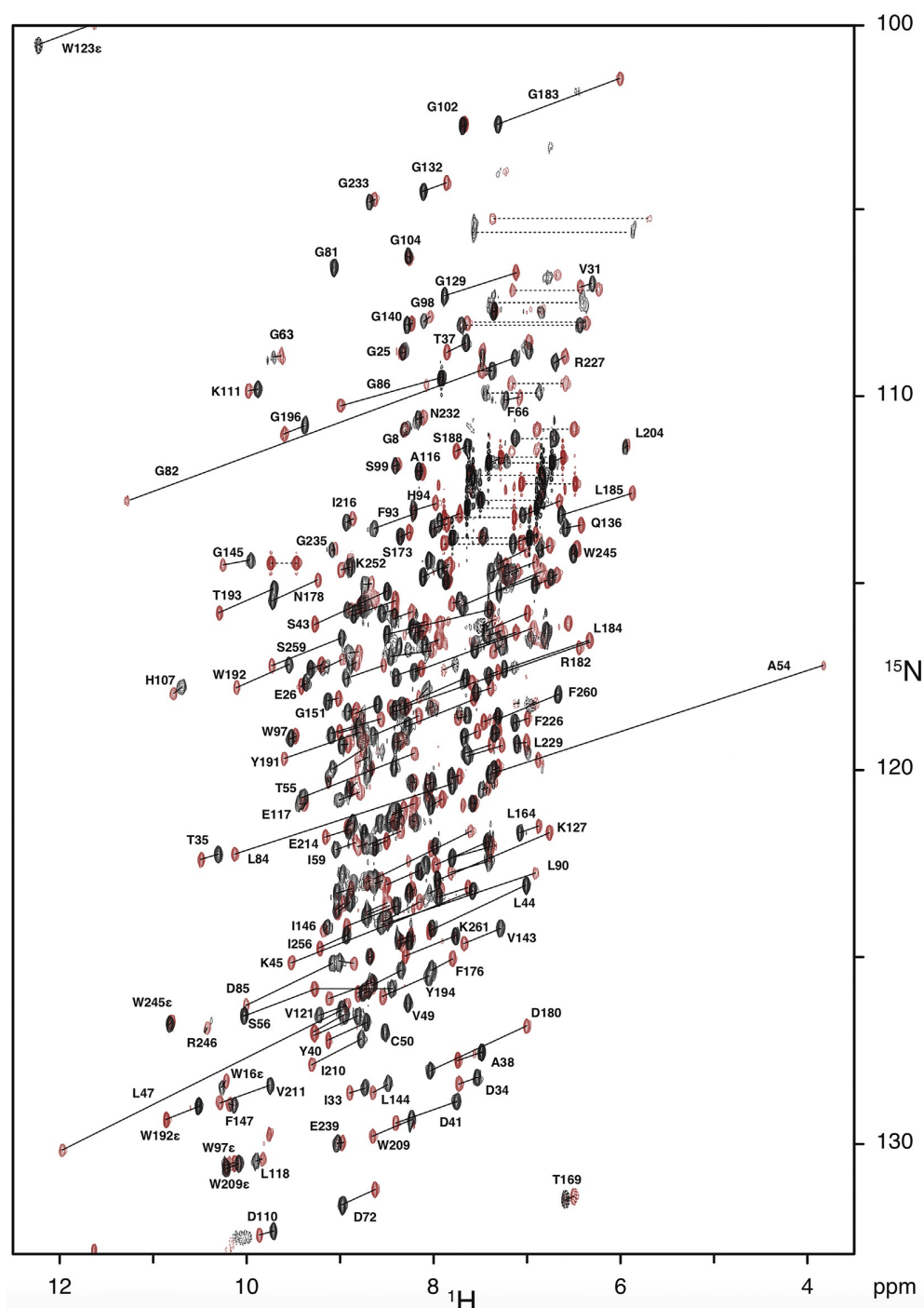


Fig. 23 Overlay of the ^1H - ^{15}N HSQC spectra of uniformly ^{15}N labeled hCA II S50C Lu-DOTA-M8-SSPy (*black*) and hCA II S50C Tm-DOTA-M8-SSPy (*red*) recorded at 600 MHz proton frequency (Aliased signals with negative intensity are shown with dashed contour lines, PCS are indicated with solid and NH_2 groups with dashed lines, water signal was omitted for clarity). Due to the highly crowded overlay of the spectra, an ambiguous result could be obtained upon direct assignment of all PCS.⁴¹ Reproduced by permission of The Royal Society of Chemistry from Zimmermann, K.; Joss, D.; Müntener, T.; Nogueira, E. S.; Schäfer, M.; Knörr, L.; Monnard, F. W.; Häussinger, D. *Chem. Sci.* **2019**, *10*, 5064–5072.

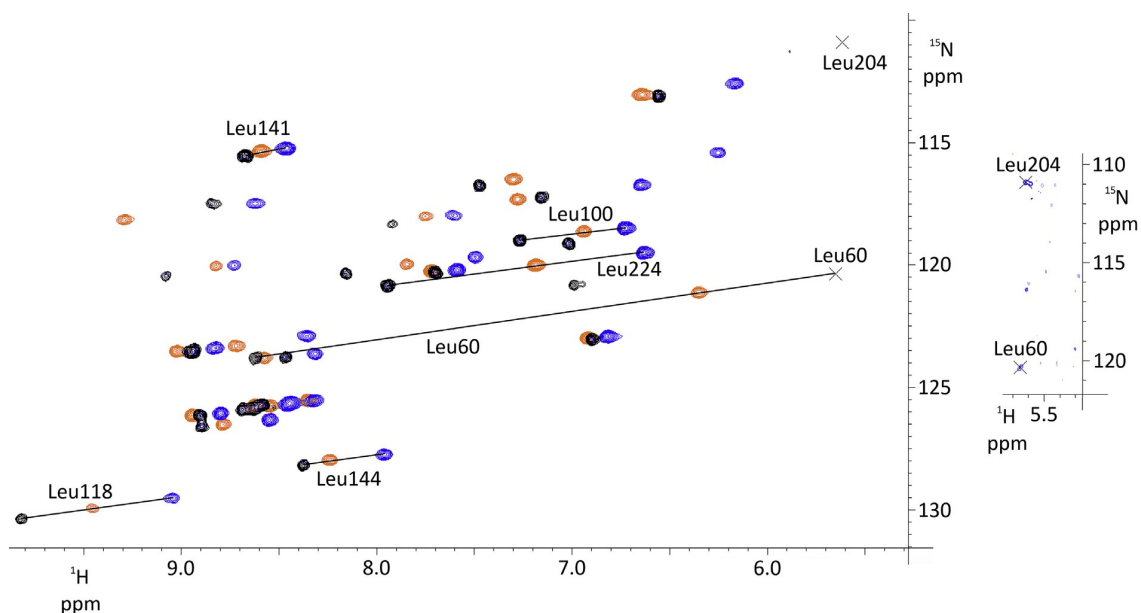


Fig. 24 Overlay of ^1H - ^{15}N HSQC spectra of selectively ^{15}N leucine labeled hCA II S166C labeled with Tm-DOTA-M8-(4*R*,4*S*)-SSPy³⁴ (orange), Tm-P4M4-DOTA (blue) and Lu-P4M4-DOTA (black) measured at 600 MHz proton frequency (left). Remarkably, due to the low spectral complexity, the PCSs can be readily assigned and leucine-derived anisotropy parameters can be fitted. Presence of peaks affected in intensity due to the applied water suppression method is demonstrated in an extra panel (right).³⁶

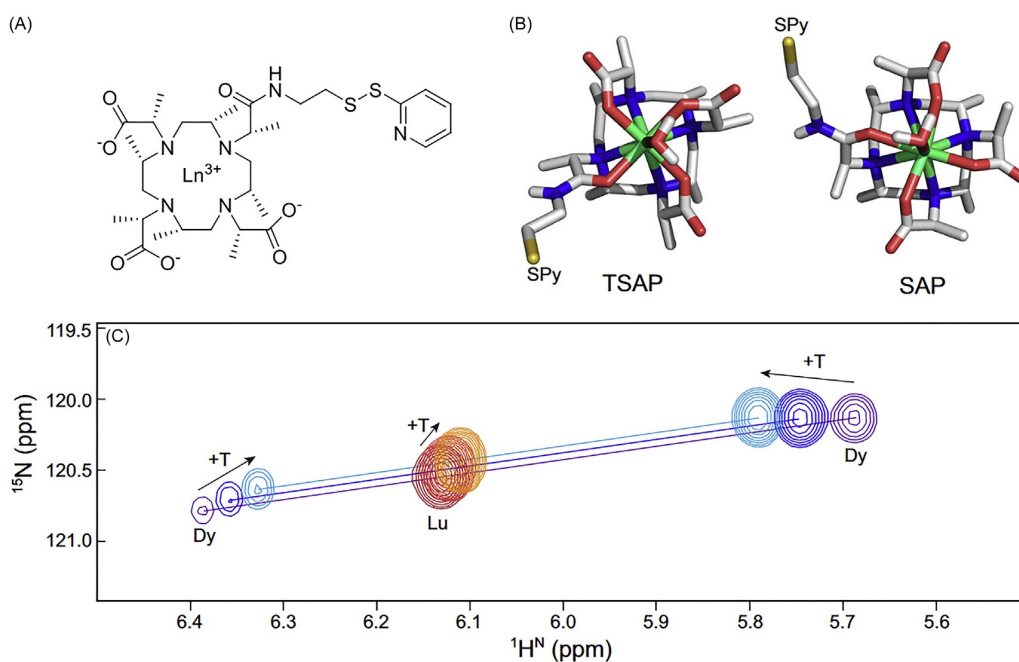


Fig. 25 Influence of different temperatures on shifts induced by the Dy-DOTA-M8-(8*S*)-SSPy LCT²⁹ on a ubiquitin S57C construct (structure of the LCT depicted in part A).⁷⁰ The major and minor species originate in two conformational isomers (B) that interconvert on a slow timescale (twisted square antiprism (TSAP) and square antiprism (SAP), see also Section 2.6). The observed temperature dependence of the induced PCS, i.e. the movement of the paramagnetic shift to the diamagnetic reference peak upon increase of the measuring temperature, can be used in order to unambiguously assign PCSs in a crowded spectral region of a ^1H - ^{15}N HSQC experiment (C, 288 K: purple, 298 K: blue, 308 K: cyan).⁷⁰ Strickland, M.; Schwieters, C. D.; Göbl, C.; Opina, A. C. L.; Strub, M.-P.; Swenson, R. E.; Vasalatiy, O.; Tjandra, N., *J. Biomol. NMR* **2016**, *66*, 125–139.

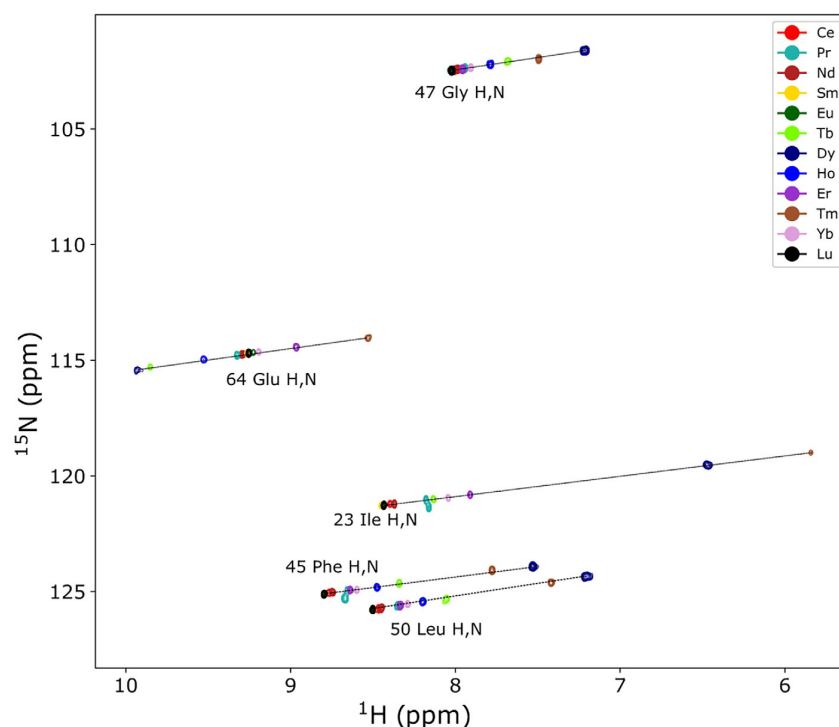


Fig. 26 Spectral overlay of a selection of peaks in ^1H - ^{15}N HSQC experiments of Ln-DOTA-M7FPy,³⁵ loaded with various lanthanoid ions, attached to an ubiquitin S57C construct. The measurement of multiple lanthanoid ions allows to critically assess the PCS assignment in crowded regions of a ^1H - ^{15}N HSQC spectrum by comparing the induced patterns of shifts.

Table 1 Anisotropy parameters for the hCA II constructs tagged with Tm-DOTA-M8-(8S)-SSPy.

Parameter	Unit	S50C Tm-DOTA	S166C Tm-DOTA	S217C Tm-DOTA	S220C Tm-DOTA	S166C Tm-Thiazole
No. of PCS	–	366	397	364	366	44
$\Delta\chi_{\text{ax}}$	$[10^{-32} \text{ m}^3]$	21.6 ± 1.2	38.5 ± 2.0	25.7 ± 1.0	23.6 ± 0.9	34.7 ± 0.6
$\Delta\chi_{\text{rh}}$	$[10^{-32} \text{ m}^3]$	8.5 ± 0.7	8.0 ± 1.0	13.2 ± 0.6	4.3 ± 0.3	13.3 ± 1.1
x	$[\text{\AA}]$	-27.8 ± 0.3	-16.3 ± 0.4	-24.9 ± 0.2	-13.1 ± 0.3	-11.8 ± 0.5
y	$[\text{\AA}]$	13.7 ± 0.3	-3.6 ± 0.4	-17.7 ± 0.3	-26.7 ± 0.3	-1.7 ± 0.1
z	$[\text{\AA}]$	18.1 ± 0.3	-11.2 ± 0.4	19.6 ± 0.2	3.2 ± 0.2	-11.0 ± 0.2
α	$[\text{^\circ}]$	104.1 ± 1.6	52.2 ± 1.8	143.7 ± 0.8	14.9 ± 1.4	119.4 ± 4.4
β	$[\text{^\circ}]$	142.3 ± 1.1	123.6 ± 1.4	70.9 ± 0.5	153.6 ± 0.6	162.1 ± 0.5
γ	$[\text{^\circ}]$	116.2 ± 1.7	140.3 ± 5.6	125.5 ± 1.0	1.0 ± 2.6	44.8 ± 3.7
Q		0.072	0.037	0.064	0.106	0.061
<i>Monte-Carlo</i> structure variation with $\sigma = 0.5 \text{ \AA}$.						
$\Delta\chi_{\text{ax}}$	$[10^{-32} \text{ m}^3]$	21.1 ± 0.9	37.4 ± 0.9	25.5 ± 1.1	23.0 ± 2.1	34.3 ± 0.8
$\Delta\chi_{\text{rh}}$	$[10^{-32} \text{ m}^3]$	8.5 ± 0.6	7.8 ± 0.6	13.1 ± 0.6	4.4 ± 0.8	12.7 ± 1.6
x	$[\text{\AA}]$	-27.5 ± 0.3	-16.2 ± 0.2	-24.8 ± 0.2	-13.0 ± 0.6	-12.0 ± 0.7
y	$[\text{\AA}]$	13.6 ± 0.2	-3.6 ± 0.2	-17.5 ± 0.3	-26.4 ± 0.6	-1.8 ± 0.2
z	$[\text{\AA}]$	18.2 ± 0.2	-11.0 ± 0.2	19.6 ± 0.2	3.2 ± 0.5	-10.8 ± 0.4
α	$[\text{^\circ}]$	104.0 ± 1.8	52.7 ± 1.2	143.3 ± 0.8	16.2 ± 3.5	121.6 ± 6.8
β	$[\text{^\circ}]$	141.8 ± 0.8	123.1 ± 0.7	71.2 ± 0.7	153.7 ± 1.5	161.6 ± 1.1
γ	$[\text{^\circ}]$	115.9 ± 1.7	141.1 ± 2.7	125.1 ± 1.1	3.9 ± 6.6	45.1 ± 5.2
Q		0.071	0.035	0.068	0.106	0.059

Monte-Carlo protocol where random subsets consisting of 20% of the available PCS were used.

Reproduced by permission of The Royal Society of Chemistry from Zimmermann, K.; Joss, D.; Müntener, T.; Nogueira, E. S.; Schäfer, M.; Knörr, L.; Monnard, F. W.; Häussinger, D. *Chem. Sci.* **2019**, *10*, 5064–5072.

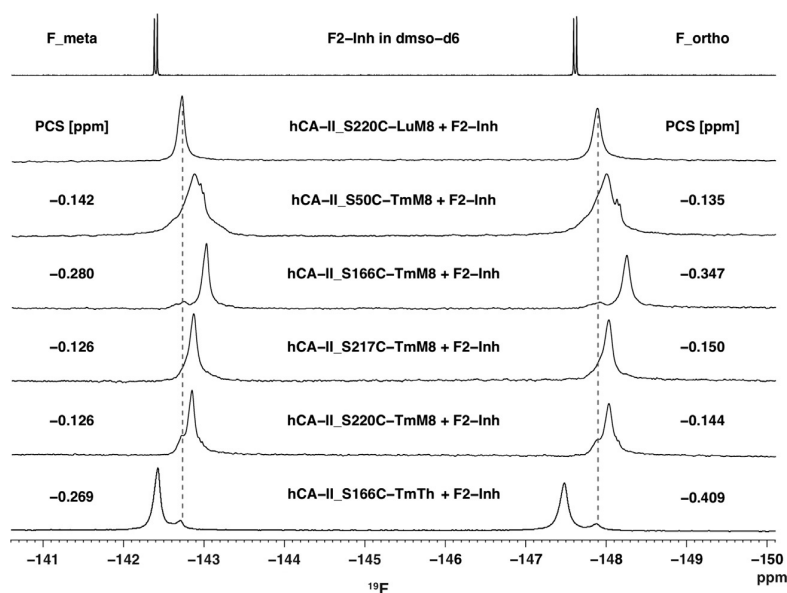


Fig. 27 One-dimensional ^{19}F NMR spectra of the difluorinated ligand F2-Inh incorporated into hCA II constructs that were labeled at different sites within the protein. The observed peaks of the ligand incorporated in the different hCA II constructs are significantly broadened when compared with the unbound ligand due to the higher rotational correlation time. Since the different hCA II constructs labeled with a diamagnetic LCT did not exhibit any differences, a hCA II S220C construct labeled with Lu-DOTA-M8-(8*S*)-SSPy was used as reference. The one-dimensional ^{19}F NMR spectra were referenced using an internal signal of trifluoroacetate at -79.0 ppm. Reproduced by permission of The Royal Society of Chemistry from Zimmermann, K.; Joss, D.; Müntener, T.; Nogueira, E. S.; Schäfer, M.; Knörr, L.; Monnard, F. W.; Häussinger, D. *Chem. Sci.* **2019**, *10*, 5064–5072.

$$\delta^{PCS} = \frac{1}{12\pi r^3} \left[\Delta\chi_{ax} (3 \cos^2 \theta - 1) + \frac{3}{2} \Delta\chi_{rh} \sin^2 \theta \cos 2\phi \right]$$

The position of the observed spin characterized by the distance to the metal center r as well as the azimuth (ϕ) and zenith (θ) angles leads then to the calculated PCS by taking into account the axial and rhombic anisotropy ($\Delta\chi_{ax}$ and $\Delta\chi_{rh}$) parameters.

In order to calculate the position of the fluorine atom(s) of the chosen sulfonamide ligands, the fitted anisotropy tensors based on the PCS of the amide N—H spin pairs can be used to back-calculate the position of the ligand by taking as input the ^{19}F PCS measured on the ligand while being complexed with the tagged protein. In order to obtain the position in an unambiguous fashion by using least-squares fitting, four different anisotropy tensors were used in the study. Furthermore, in order to further enhance the accuracy, a fifth anisotropy tensor with an additional orthogonal isosurface can be incorporated during the least-squares minimization. The square residual that is minimized during the process is given by the following definition:

$$s(x, y, z) = \sum_{i=1}^n \left(\delta_i^{PCS}(x'_i, y'_i, z'_i) - \delta_i^{PCS}(\text{experiment}) \right)^2$$

The definition is composed by the sum of the deviation between the experimentally measured PCS and the back-calculated PCS, whereas i is the corresponding tensor index and $\delta_i^{PCS}(\text{experiment})$ is the experimental PCS. The back-calculated PCS is given by the semi-empirical formula of Kurland and McGarvey¹⁷⁸ for the PCS in Cartesian coordinates that incorporates the positioning of the spin (x, y, z), the distance from the spin to the paramagnetic center as well as the anisotropy parameters $\Delta\chi_{ax,i}$ and $\Delta\chi_{rh,i}$ which give the axiality and rhombicity of the earlier determined anisotropy tensor.

$$\delta_i^{PCS}(x'_i, y'_i, z'_i) = \frac{1}{12\pi r_i^3} \left[\Delta\chi_{ax,i} \frac{2(z'_i)^2 - (x'_i)^2 - (y'_i)^2}{r_i^2} + \frac{3}{2} \Delta\chi_{rh,i} \frac{(x'_i)^2 - (y'_i)^2}{r_i^2} \right]$$

The difference between back-calculated and measured PCS, i.e. the function $s(x, y, z)$, leads then straightforward to the position of the fluorine nuclei of the tested ligand. However, one has further to account for the uncertainties of the used anisotropy tensors. Therefore, a Monte-Carlo protocol was implemented in the algorithm that introduces errors based on a random seed and performs 10,000 iterations in order to have a statistically valid sampling number.¹⁷⁹ The introduced uncertainties are based on the errors of the tensor values determined during the fitting process using Numbat.⁹⁷

By application of the described protocol in our model system, the tested inhibitors were successfully localized using PCS restraints with an accuracy of up to 0.8 \AA . Furthermore, the inhibitors were localized over a through-space distance of $22\text{--}38 \text{ \AA}$, a previously unprecedented distance range. Earlier studies reported localizations of ligands over distances of $10\text{--}25 \text{ \AA}$.^{40,102,151} results

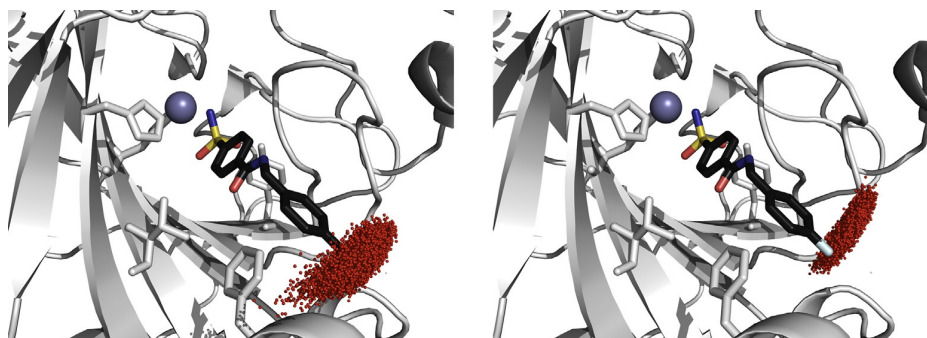


Fig. 28 Point cloud obtained in the Monte-Carlo calculation of the position of the fluorine atoms of FM-520 (*left*) and FM-519 (*right*). Light blue sticks indicate fluorine atoms, while the blue sphere represents the zinc(II) ion (modified PDB 1G54).^{41,177} Reproduced by permission of The Royal Society of Chemistry from Zimmermann, K.; Joss, D.; Müntener, T.; Nogueira, E. S.; Schäfer, M.; Knörr, L.; Monnard, F. W.; Häussinger, D. *Chem. Sci.* **2019**, *10*, 5064–5072.

that were already considered as highly promising given that for NOESY a distance limitation of $\sim 5\text{--}8$ Å applies and a complete sidechain assignment is required. The key factor for the distance range that was monitored in the present study lies in the use of the strongly paramagnetic LCTs that lead to large PCS detected for the ligand atoms. With the newest generations of strongly paramagnetic LCTs available today, distances over 200 Å can be sampled in the optimal case, i.e. given a favorable alignment of the induced anisotropy with the protein.¹⁸ For proteins with a mainly globular fold, the newest generation of LCTs would even allow to localize ligands in proteins over 500 kDa, which includes most proteins suitable for NMR spectroscopy by today.

The determined position for the ligands FM-519 and FM-520 shows deviations of 3.3 and 0.8 Å when compared to the X-ray structure of hCA II incorporating a pentafluorobenzyl-substituted ligand (PDB 1G54)¹⁷⁷ (Fig. 28). Interestingly, the found deviations of the position found for the ligand F2-Inh by PCS and in the X-ray structure (PDB 1G52)¹⁷⁷ deviate more strongly, although the intramolecular distance between the two fluorine atoms in *ortho*- and *meta* position of the phenyl ring of the ligand is very accurately reproduced (determined as 3.5 Å compared to 2.8 Å in the X-ray structure (modified PDB 1G52)¹⁷⁷). However, based on our results and MP2 calculations that investigate the interactions of differently fluorinated benzyl rings with another nonfluorinated benzyl ring (Motif 1b on page 2 of Ref. ¹⁸⁰), it was proposed that the outer aromatic ring of the ligand aligns in solution different from the conformation found in the crystal structure due to the interactions with the neighboring phenyl substituent of Phe-131. By taking into account the possible structural changes of the outer phenyl ring of the ligand in solution, the positional deviations coincided within 1.6 Å with the fluorine in *meta* position of the difluorophenyl substituent and 2.6 Å for the fluorine in *ortho* position.

4.5 Determination of the Ligand Position by Analysis of the Isosurfaces of the Anisotropy Parameters

In principle, the position of the investigated ligand can very simply be obtained by using only the intersection of three back-calculated isosurfaces for a specific PCS based on the respective previously determined anisotropy tensors. This approach wins clearly over the Monte-Carlo method, which incorporates more than three anisotropy tensors, by its pure straightforwardness. However, by taking in account only three anisotropy tensors, the result could be corroborated due to an unfavorable alignment and thus not sufficient orthogonality of the used isosurfaces. In order to investigate the described approach and quantify its deviations and pitfalls, the position was determined for all three applied ligands in the study with all possible combinations of tensor isosurfaces. The orthogonality of tensors was determined by the intersection angle of the normal vectors at the intersection point.

The obtained results clearly show that the closer the intersection angles match 90 degrees, the more precise the position can be determined. In order to reflect the found intersection angles and the suitability of a given combination of isosurfaces, an angle score was defined by calculating the sum of the deviations of the intersection angles to 90 degrees and dividing the result by 3. Therefore, an angle score of 0 degree states perfect orthogonality of the used isosurfaces, while 90 degrees indicates parallel isosurfaces. We found that the combination of isosurfaces with an angle score below 30 degrees closely match the positions determined using the Monte-Carlo protocol (average deviation for combinations of isosurfaces with an angle score below 30 degrees of all four ligands: 2.3 Å) (Fig. 29). For combinations exhibiting an angle score of over 40 degrees, the deviations amount to 7 Å or more. Therefore, it can be concluded that the use of only three different isosurfaces can be used to localize the ligand given that the used isosurfaces show a favorable angle score. Even for combinations of isosurfaces with an unfavorable angle score, the achieved accuracy would still allow a coarse-grained localization of the ligand and thus enable to discriminate two different, distant binding sites for the ligands within the investigated protein.

4.6 ¹H-¹⁹F-HOESY Experiments for Determination of ¹H PCS of the Ligand and Protein–Ligand Interactions

Since it is impossible to discriminate the proton signals of the bound ligand and the protein host present in equimolar ratios within the sample, ¹H-¹⁹F heteronuclear Overhauser effect spectroscopy (HOESY) experiments were measured in order to determine the ¹H

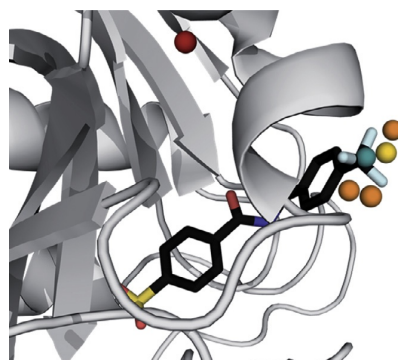


Fig. 29 Points of intersection of isosurfaces induced by LCTs on different tagging sites (ligand FM-520, center of gravity of the CF_3 fluorine atoms: cyan (modified PDB 1G54),¹⁷⁷ fluorine atoms: light blue sticks, position obtained by least square minimization using all four tensors: gold, positions obtained by intersection of three isosurfaces in all possible combinations with an angle score below 30: orange, positions obtained by intersection of three isosurfaces in all possible combinations with an angle score above 30: red).⁴¹ Reproduced by permission of The Royal Society of Chemistry from Zimmermann, K.; Joss, D.; Müntener, T.; Nogueira, E. S.; Schäfer, M.; Knörr, L.; Monnard, F. W.; Häussinger, D. *Chem. Sci.* **2019**, *10*, 5064–5072.

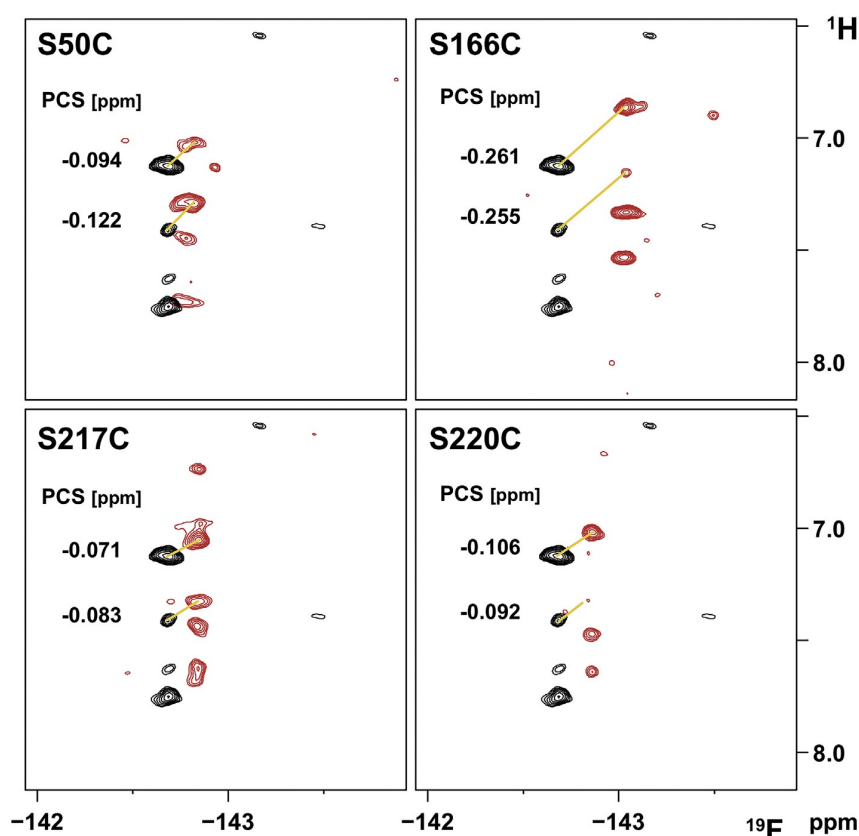


Fig. 30 HOESY spectra of hCA II constructs incorporating the ligand F2-Inh. The correlation peaks represent through-space interactions of ^{19}F with ^1H . Due to the narrow sweep width, the signals for aliphatic protons appear at their aliased frequency.¹⁸¹ Zimmermann, K.; Joss, D.; Müntener, T.; Nogueira, E. S.; Schäfer, M.; Knörr, L.; Monnard, F. W.; Häussinger, D., *Chem. Sci.* **2019**, *10*, 5064–5072.

PCS of proton atoms of the bound ligand (F2-Inh), which are close in space to the fluorine atoms. The experiments revealed four correlation peaks with variable intensities (Fig. 30). Due to the small sweep width, the peaks appear at their aliased frequency. In order to analyze their original NMR shift, spectra with a larger spectral window were recorded. The obtained results show that three peaks originate from correlation of fluorine with sidechain methyl groups, since their proton frequency was found in the aliphatic region between 0.4 and 0.75 ppm. The strongest ^1H - ^{19}F HOESY correlation peak was considered to constitute the cross peak of the proton in *para*-position of the fluorinated phenyl ring of the ligand with the fluorine in *ortho*- and *meta* position.

Upon determination of the ^1H PCS, the aromatic proton atom of F2-Inh in *para* position was localized within the protein by fitting of its experimentally determined PCS to the back-calculated PCS based on the anisotropy tensors derived from the H—N signals. The positional deviation of the proton, based on the PCS approach, amounted to only 1.4 Å when compared with the X-ray structure of hCA II incorporating the difluorinated ligand F2-Inh (PDB 1G52)¹⁷⁷. Interactions of the benzylic fluorine atoms with side-chain methyl groups are clearly observed and the resonance at 0.41 ppm was, based on its observed PCS, attributed to interactions with the methyl groups of the residue Val-135.

5 Conclusion and Outlook

To conclude, the strongly paramagnetic LCTs that are available today yield accurate restraints for the structural elucidation of biomacromolecules in solution, i.e. *in vitro* as well as *in cellulo*. PCS analysis offers the possibility to investigate the three-dimensional structure, conformations and dynamics of biomacromolecules. Furthermore, the site of ligand-binding and the structure of the protein–ligand complex can be resolved by the structural restraints obtained by application of an LCT.

As demonstrated by various research groups, nuclei as ^1H , ^{13}C and ^{19}F in combination with different NMR approaches, e.g. conventional ^1H or ^{19}F one-dimensional spectra, ^{19}F CEST and ^1H relaxation dispersion, can be applied in order to achieve a reasonable signal discrimination of the ligand signals from the protein. The chosen approach depends on the measured nucleus as well as on the affinity of the ligand, i.e. ligands that show weak, intermediate or strong binding. The distance range that can be covered according to the reported literature (10–38 Å) renders this unique method highly suitable for biologically relevant applications. This can be achieved by the application of covalently binding and strongly paramagnetic LCTs to the protein that generate large PCSs and provide therefore accurate structural restraints for the measured nuclei.

By use of human carbonic anhydrase II and incorporated sulfonamide ligands, which interact with nM to pM K_d -values with the zinc metal center, it was shown that it is possible to unambiguously localize a ligand bound ^{19}F atom within a protein approaching an accuracy of 0.8 Å. Solely based on the fluorine PCS caused by LCTs over a distance of 22–38 Å by measuring 4–5 one-dimensional ^{19}F spectra per ligand, one unique solution for the position of the fluorine atoms was obtained. Besides Monte-Carlo calculations with 4 or 5 different anisotropy tensors, graphical analysis using only three isosurfaces shows a high accuracy as long as sufficient orthogonality of the isosurfaces is confirmed by application of an angle score. When only combinations are taken into account showing an angle score below 30 degrees, the obtained accuracy allows a clear localization of the fluorine atom and the position matches the one determined in Monte-Carlo protocols using four tensors. By omitting the need for solvent suppression, chemical modification of the ligand and extensive measurement times during the screening process, the method constitutes a fast, reliable and convenient approach to screen a high number of fluorine-containing ligands for a specific protein.

The use of LCTs for the characterization of protein–ligand complexes in solution can deliver accurate structural restraints over a large distance and thus holds great promise for future applications in structural biology and rational drug design. With the newest generation of LCTs in hand, the localization of ligands by PCS can further be extended to proteins with molecular weights significantly larger than 30 kDa and large, multimeric protein complexes.

Acknowledgment

The Chemistry Department of the University of Basel and the Swiss National Science Foundation grant 200021_130263 are acknowledged for financial support. Biological structures were generated using the open source software PyMOL (<http://www.pymol.org/>).

References

1. Wüthrich, K. *Angew. Chem. Int. Ed. Engl.* **2003**, *42*, 3340–3363.
2. Pervushin, K.; Riek, R.; Wider, G.; Wüthrich, K. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 12366–12371.
3. Salzmann, M.; Pervushin, K.; Wider, G.; Senn, H.; Wüthrich, K. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 13585–13590.
4. Salzmann, M.; Wider, G.; Pervushin, K.; Senn, H.; Wüthrich, K. *J. Am. Chem. Soc.* **1999**, *121*, 844–848.
5. Salzmann, M.; Wider, G.; Pervushin, K.; Wüthrich, K. *J. Biomol. NMR* **1999**, *15*, 181–184.
6. Kumar, A.; Ernst, R. R.; Wüthrich, K. *Biochem. Biophys. Res. Commun.* **1980**, *95*, 1–6.
7. Wider, G.; Macura, S.; Kumar, A.; Ernst, R. R.; Wüthrich, K. *J. Magn. Reson.* **1984**, *56*, 207–234.
8. Ujwal, R.; Cascio, D.; Colletier, J.-P.; Faham, S.; Zhang, J.; Toro, L.; Ping, P.; Abramson, J. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 17742.
9. Böhm, R.; Amodeo, G. F.; Murlidaran, S.; Chavali, S.; Wagner, G.; Winterhalter, M.; Brannigan, G.; Hiller, S. *Structure* **2019**. <https://doi.org/10.1016/j.str.2019.11.015>.
10. Wüthrich, K. *Nat. Struct. Biol.* **2001**, *8*, 923–925.
11. Marion, D. *Mol. Cell. Proteomics* **2013**, *12*, 3006.
12. Zhu, G.; Xia, Y.; Nicholson, L. K.; Sze, K. H. *J. Magn. Reson.* **2000**, *143*, 423–426.
13. Renner, C.; Schleicher, M.; Moroder, L.; Holak, T. A. *J. Biomol. NMR* **2002**, *23*, 23–33.
14. Wijmenga, S. S.; van Buuren, B. N. M. *Prog. Nucl. Magn. Reson. Spectrosc.* **1998**, *32*, 287–387.
15. Jamin, N.; Toma, F. *Prog. Nucl. Magn. Reson. Spectrosc.* **2001**, *38*, 83–114.
16. Fürtig, B.; Richter, C.; Wöhnert, J.; Schwalbe, H. *ChemBioChem* **2003**, *4*, 936–962.
17. Al-Hashimi, H. M. *J. Magn. Reson.* **2013**, *237*, 191–204.

18. Joss, D.; Häussinger, D. *Prog. Nucl. Magn. Reson. Spectrosc.* **2019**, *114–115*, 284–312.
19. Barry, C. D.; North, A. C. T.; Glasel, J. A.; Williams, R. J. P.; Xavier, A. V. *Nature* **1971**, *232*, 236–245.
20. Bertini, I.; Turano, P.; Vila, A. J. *Chem. Rev.* **1993**, *93*, 2833–2932.
21. Banci, L.; Bertini, I.; Bren, K. L.; Cremonini, M. A.; Gray, H. B.; Luchinat, C.; Turano, P. *J. Biol. Inorg. Chem.* **1996**, *1*, 117–126.
22. Banci, L.; Bertini, I.; Savellini, G. G.; Romagnoli, A.; Turano, P.; Cremonini, M. A.; Luchinat, C.; Gray, H. B. *Proteins Struct. Funct. Bioinf.* **1997**, *29*, 68–76.
23. Bertini, I.; Luchinat, C.; Parigi, G.; Pierattelli, R. *ChemBioChem* **2005**, *6*, 1536–1549.
24. Wöhnert, J.; Franz, K. J.; Nitz, M.; Imperiali, B.; Schwalbe, H. *J. Am. Chem. Soc.* **2003**, *125*, 13338–13339.
25. Franz, K. J.; Nitz, M.; Imperiali, B. *ChemBioChem* **2003**, *4*, 265–271.
26. Nitz, M.; Franz, K. J.; Maglathlin, R. L.; Imperiali, B. *ChemBioChem* **2003**, *4*, 272–276.
27. Saio, T.; Ogura, K.; Yokochi, M.; Kobashigawa, Y.; Inagaki, F. *J. Biomol. NMR* **2009**, *44*, 157–166.
28. Keizers, P. H.; Saragliadis, A.; Hiruma, Y.; Overhand, M.; Ubbink, M. *J. Am. Chem. Soc.* **2008**, *130*, 14802–14812.
29. Häussinger, D.; Huang, J.-R.; Grzesiek, S. *J. Am. Chem. Soc.* **2009**, *131*, 14761–14767.
30. Peters, F.; Maestre-Martinez, M.; Leonov, A.; Kovacic, L.; Becker, S.; Boelens, R.; Griesinger, C. *J. Biomol. NMR* **2011**, *51*, 329–337.
31. Graham, B.; Loh, C. T.; Swarbrick, J. D.; Ung, P.; Shin, J.; Yagi, H.; Jia, X.; Chhabra, S.; Barlow, N.; Pintacuda, G.; Huber, T.; Otting, G. *Bioconjug. Chem.* **2011**, *22*, 2118–2125.
32. Yang, F.; Wang, X.; Pan, B.-B.; Su, X.-C. *Chem. Commun.* **2016**, *52*, 11535–11538.
33. Müntener, T.; Häussinger, D.; Selenko, P.; Theillet, F.-X. *J. Phys. Chem. Lett.* **2016**, *7*, 2821–2825.
34. Joss, D.; Walliser, R. M.; Zimmermann, K.; Häussinger, D. *J. Biomol. NMR* **2018**, 29–38.
35. Müntener, T.; Kottelat, J.; Huber, A.; Häussinger, D. *Bioconjug. Chem.* **2018**, *29*, 3344–3351.
36. Joss, D.; Bertrams, M.-S.; Häussinger, D. *Chem. Eur. J.* **2019**, *25*, 11910–11917.
37. Joss, D.; Häussinger, D. *Chem. Commun.* **2019**, *55*, 10543–10546.
38. Bertrand, J. A.; Auger, G.; Martin, L.; Fanchon, E.; Blanot, D.; Le Beller, D.; van Heijenoort, J.; Dideberg, O. *J. Mol. Biol.* **1999**, *289*, 579–590.
39. Saio, T.; Ogura, K.; Kumeta, H.; Kobashigawa, Y.; Shimizu, K.; Yokochi, M.; Kodama, K.; Yamaguchi, H.; Tsujishita, H.; Inagaki, F. *Sci. Rep.* **2015**, *5*, 16685.
40. Guan, J.-Y.; Keizers, P. H. J.; Liu, W.-M.; Löhr, F.; Skinner, S. P.; Heeneman, E. A.; Schwalbe, H.; Ubbink, M.; Siegal, G. *J. Am. Chem. Soc.* **2013**, *135*, 5859–5868.
41. Zimmermann, K.; Joss, D.; Müntener, T.; Nogueira, E. S.; Schäfer, M.; Knörr, L.; Monnard, F. W.; Häussinger, D. *Chem. Sci.* **2019**, *10*, 5064–5072.
42. Strickland, M.; Catazaro, J.; Rajasekaran, R.; Strub, M.-P.; O'Hern, C.; Bermejo, G. A.; Summers, M. F.; Marchant, J.; Tjandra, N. *J. Am. Chem. Soc.* **2019**, *141*, 1430–1434.
43. Göbl, C.; Resch, M.; Strickland, M.; Hartmüller, C.; Viertler, M.; Tjandra, N.; Madl, T. *Angew. Chem. Int. Ed. Engl.* **2016**, *55*, 14847–14851.
44. Brewer, K. D.; Bacaj, T.; Cavalli, A.; Camilloni, C.; Swarbrick, J. D.; Liu, J.; Zhou, A.; Zhou, P.; Barlow, N.; Xu, J.; Seven, A. B.; Prinslow, E. A.; Voletí, R.; Häussinger, D.; Bonvin, A. M. J. J.; Tomchick, D. R.; Vendruscolo, M.; Graham, B.; Südhof, T. C.; Rizo, J. *Nat. Struct. Mol. Biol.* **2015**, *22*, 555–564.
45. Canales, A.; Boos, I.; Perkams, L.; Karst, L.; Lubert, T.; Karagiannis, T.; Dominguez, G.; Cañada, F. J.; Pérez-Castells, J.; Häussinger, D.; Unverzagt, C.; Jiménez-Barbero, J. *Angew. Chem. Int. Ed. Engl.* **2017**, *56*, 14987–14991.
46. Rodríguez-Castañeda, F.; Haber, P.; Leonov, A.; Griesinger, C. *Magn. Reson. Chem.* **2006**, *44*, S10–S16.
47. Pintacuda, G.; John, M.; Su, X.-C.; Otting, G. *Acc. Chem. Res.* **2007**, *40*, 206–212.
48. John, M.; Otting, G. *ChemPhysChem* **2007**, *8*, 2309–2313.
49. Otting, G. *J. Biomol. NMR* **2008**, *42*, 1–9.
50. Otting, G. *Annu. Rev. Biophys.* **2010**, *39*, 387–405.
51. Su, X.-C.; Otting, G. *J. Biomol. NMR* **2010**, *46*, 101–112.
52. Allen, K. N.; Imperiali, B. *Curr. Opin. Chem. Biol.* **2010**, *14*, 247–254.
53. Koehler, J.; Meiler, J. *Prog. Nucl. Magn. Reson. Spectrosc.* **2011**, *59*, 360–389.
54. Keizers, P. H. J.; Ubbink, M. *Prog. Nucl. Magn. Reson. Spectrosc.* **2011**, *58*, 88–96.
55. Bhaumik, A.; Luchinat, C.; Parigi, G.; Ravera, E.; Rinaldelli, M. *CrystEngComm* **2013**, *15*, 8639–8656.
56. Hass, M. A. S.; Ubbink, M. *Curr. Opin. Struct. Biol.* **2014**, *24*, 45–53.
57. Prestegard, J. H.; Agard, D. A.; Moremen, K. W.; Lavery, L. A.; Morris, L. C.; Pederson, K. J. *Magn. Reson.* **2014**, *241*, 32–40.
58. Göbl, C.; Madl, T.; Simon, B.; Sattler, M. *Prog. Nucl. Magn. Reson. Spectrosc.* **2014**, *80*, 26–63.
59. Liu, W.-M.; Overhand, M.; Ubbink, M. *Coord. Chem. Rev.* **2014**, *273–274*, 2–12.
60. Kato, K.; Yamaguchi, T. *Glycoconjugate J.* **2015**, *32*, 505–513.
61. Carlon, A.; Ravera, E.; Andralojć, W.; Parigi, G.; Murshudov, G. N.; Luchinat, C. *Prog. Nucl. Magn. Reson. Spectrosc.* **2016**, *92–93*, 54–70.
62. Nitsche, C.; Otting, G. *Prog. Nucl. Magn. Reson. Spectrosc.* **2017**, *98–99*, 20–49.
63. Su, X.-C.; Chen, J.-L. *Acc. Chem. Res.* **2019**, *52*, 1675–1686.
64. Saio, T.; Ishimori, K. *Biochim. Biophys. Acta. Gen. Subj.* **2020**. <https://doi.org/10.1016/j.bbagen.2019.03.018>. in press.
65. Su, X.-C.; Man, B.; Beeren, S.; Liang, H.; Simonsen, S.; Schmitz, C.; Huber, T.; Messerle, B. A.; Otting, G. *J. Am. Chem. Soc.* **2008**, *130*, 10486–10487.
66. Su, X.-C.; Liang, H.; Loscha, K. V.; Otting, G. *J. Am. Chem. Soc.* **2009**, *131*, 10352–10353.
67. Loh, C. T.; Ozawa, K.; Tuck, K. L.; Barlow, N.; Huber, T.; Otting, G.; Graham, B. *Bioconjug. Chem.* **2013**, *24*, 260–268.
68. Lee, M. D.; Loh, C. T.; Shin, J.; Chhabra, S.; Dennis, M. L.; Otting, G.; Swarbrick, J. D.; Graham, B. *Chem. Sci.* **2015**, *6*, 2614–2624.
69. Pan, B.-B.; Yang, F.; Ye, Y.; Wu, Q.; Li, C.; Huber, T.; Su, X.-C. *Chem. Commun.* **2016**, *52*, 10237–10240.
70. Strickland, M.; Schwieters, C. D.; Göbl, C.; Opina, A. C. L.; Strub, M.-P.; Swenson, R. E.; Vasalatiy, O.; Tjandra, N. *J. Biomol. NMR* **2016**, *66*, 125–139.
71. Pearce, B. J. G.; Jabar, S.; Loh, C.-T.; Szabo, M.; Graham, B.; Otting, G. *J. Biomol. NMR* **2017**, *68*, 19–32.
72. Wu, Z.; Lee, M. D.; Carruthers, T. J.; Szabo, M.; Dennis, M. L.; Swarbrick, J. D.; Graham, B.; Otting, G. *Bioconjug. Chem.* **2017**, *28*, 1741–1748.
73. Barbieri, R.; Bertini, I.; Cavallaro, G.; Lee, Y.-M.; Luchinat, C.; Rosato, A. *J. Am. Chem. Soc.* **2002**, *124*, 5581–5587.
74. Kamen, D. E.; Cahill, S. M.; Girvin, M. E. *J. Am. Chem. Soc.* **2007**, *129*, 1846–1847.
75. Xu, X.; Keizers, P. H. J.; Reinle, W.; Hannemann, F.; Bernhardt, R.; Ubbink, M. *J. Biomol. NMR* **2009**, *43*, 247–254.
76. Chiliveri, S. C.; Louis, J. M.; Ghirlando, R.; Baber, J. L.; Bax, A. *J. Am. Chem. Soc.* **2018**, *140*, 34–37.
77. Barnes, C. A.; Shen, Y.; Ying, J.; Takagi, Y.; Torchia, D. A.; Sellers, J. R.; Bax, A. *J. Am. Chem. Soc.* **2019**, *141*, 9004–9017.
78. Gillespie, J. R.; Shortle, D. *J. Mol. Biol.* **1997**, *268*, 170–184.
79. Donaldson, L. W.; Skrynnykov, N. R.; Choy, W.-Y.; Muhandiram, D. R.; Sarkar, B.; Forman-Kay, J. D.; Kay, L. E. *J. Am. Chem. Soc.* **2001**, *123*, 9843–9847.
80. Iwahara, J.; Schwieters, C. D.; Clore, G. M. *J. Am. Chem. Soc.* **2004**, *126*, 5879–5896.
81. Liang, B.; Bushweller, J. H.; Tamm, L. K. *J. Am. Chem. Soc.* **2006**, *128*, 4389–4397.
82. Iwahara, J.; Clore, G. M. *Nature* **2006**, *440*, 1227–1230.
83. Iwahara, J.; Tang, C.; Marius Clore, G. *J. Magn. Reson.* **2007**, *184*, 185–195.
84. Clore, G. M.; Iwahara, J. *Chem. Rev.* **2009**, *109*, 4108–4139.
85. Gansow, O. A.; Loeffler, P. A.; Davis, R. E.; Lenkinski, R. E.; Wilcott, M. R. *J. Am. Chem. Soc.* **1976**, *98*, 4250–4258.
86. Sherry, A. D.; Ren, J.; Huskens, J.; Brücher, E.; Tóth, É.; Geraldes, C. F. C. G.; Castro, M. M. C. A.; Cacheris, W. P. *Inorg. Chem.* **1996**, *35*, 4604–4612.
87. Fonda, K. K.; Vallarino, L. M. *Inorg. Chim. Acta* **2002**, *334*, 403–410.

88. Di Pietro, S.; Piano, S. L.; Di Bari, L. *Coord. Chem. Rev.* **2011**, *255*, 2810–2820.
89. Cornilescu, G.; Bax, A. J. *Am. Chem. Soc.* **2000**, *122*, 10143–10154.
90. John, M.; Park, A. Y.; Pintacuda, G.; Dixon, N. E.; Otting, G. *J. Am. Chem. Soc.* **2005**, *127*, 17190–17191.
91. Orton, H. W.; Kuprov, I.; Loh, C.-T.; Otting, G. *J. Phys. Chem. Lett.* **2016**, *7*, 4815–4818.
92. Angyal, S. J.; Greeves, D.; Pickles, V. A. *J. Chem. Soc. Chem. Commun.* **1974**, 589–590.
93. Rodríguez-Rodríguez, A.; Esteban-Gómez, D.; de Blas, A.; Rodríguez-Blas, T.; Botta, M.; Tripiet, R.; Platas-Iglesias, C. *Inorg. Chem.* **2012**, *51*, 13419–13429.
94. Chen, K.; Tjandra, N. *Top. Curr. Chem.* **2012**, *326*, 47–67.
95. Hikone, Y.; Hirai, G.; Mishima, M.; Inomata, K.; Ikeya, T.; Arai, S.; Shirakawa, M.; Sodeoka, M.; Ito, Y. *J. Biomol. NMR* **2016**, *66*, 99–110.
96. Bertini, I.; Luchinat, C.; Parigi, G. *Prog. Nucl. Magn. Reson. Spectrosc.* **2002**, *40*, 249–273.
97. Schmitz, C.; Stanton-Cook, M. J.; Su, X. C.; Otting, G.; Huber, T. *J. Biomol. NMR* **2008**, *41*, 179–189.
98. Orton, H. W.; Huber, T.; Otting, G. *Magn. Reson. Discuss.* **2019**, 1–18.
99. Saio, T.; Yokochi, M.; Kumeta, H.; Inagaki, F. *J. Biomol. NMR* **2010**, *46*, 271–280.
100. Lescanne, M.; Skinner, S. P.; Blok, A.; Timmer, M.; Cerofolini, L.; Fragai, M.; Luchinat, C.; Ubbink, M. *J. Biomol. NMR* **2017**, *69*, 183–195.
101. Lescanne, M.; Ahuja, P.; Blok, A.; Timmer, M.; Akerud, T.; Ubbink, M. *J. Biomol. NMR* **2018**, *71*, 275–285.
102. Saio, T.; Ogura, K.; Shimizu, K.; Yokochi, M.; Burke, T. R.; Inagaki, F. *J. Biomol. NMR* **2011**, *51*, 395–408.
103. Shannon, R. D. *Acta Crystallogr. A* **1976**, *32*, 751–767.
104. Bertini, I.; Donaire, A.; Jimenez, B.; Luchinat, C.; Parigi, G.; Piccioli, M.; Poggi, L. *J. Biomol. NMR* **2001**, *21*, 85–98.
105. Bertini, I.; Janik, M. B. L.; Lee, Y.-M.; Luchinat, C.; Rosato, A. *J. Am. Chem. Soc.* **2001**, *123*, 4181–4188.
106. Mironov, V. S.; Galymetdinov, Y. G.; Ceulemans, A.; Görler-Walrand, C.; Binneemans, K. *Chem. Phys. Lett.* **2001**, *345*, 132–140.
107. Funk, A. M.; Finney, K.-L. N. A.; Harvey, P.; Kenwright, A. M.; Neil, E. R.; Rogers, N. J.; Kanthi Senanayake, P.; Parker, D. *Chem. Sci.* **2015**, *6*, 1655–1662.
108. Mason, K.; Rogers, N. J.; Suturina, E. A.; Kuprov, I.; Aguilar, J. A.; Batsanov, A. S.; Yufit, D. S.; Parker, D. *Inorg. Chem.* **2017**, *56*, 4028–4038.
109. Suturina, E. A.; Mason, K.; Gerales, C. F. G. C.; Kuprov, I.; Parker, D. *Angew. Chem. Int. Ed. Engl.* **2017**, *56*, 12215–12218.
110. Harnden, A. C.; Suturina, E. A.; Batsanov, A. S.; Senanayake, P. K.; Fox, M. A.; Mason, K.; Vonci, M.; McInnes, E. J. L.; Chilton, N. F.; Parker, D. *Angew. Chem. Int. Ed. Engl.* **2019**, *58*, 10290–10294.
111. Ma, C.; Opella, S. J. *J. Magn. Reson.* **2000**, *146*, 381–384.
112. Nitz, M.; Sherawat, M.; Franz, K. J.; Peisach, E.; Allen, K. N.; Imperiali, B. *Angew. Chem. Int. Ed. Engl.* **2004**, *43*, 3682–3685.
113. Su, X.-C.; Huber, T.; Dixon, N. E.; Otting, G. *ChemBioChem* **2006**, *7*, 1599–1604.
114. Su, X.-C.; McAndrew, K.; Huber, T.; Otting, G. *J. Am. Chem. Soc.* **2008**, *130*, 1681–1687.
115. Ikegami, T.; Verdier, L.; Sakhail, P.; Grimme, S.; Pescatore, B.; Saxena, K.; Fiebig, K. M.; Griesinger, C. *J. Biomol. NMR* **2004**, *29*, 339–349.
116. Pintacuda, G.; Moshref, A.; Leonchiks, A.; Sharipo, A.; Otting, G. *J. Biomol. NMR* **2004**, *29*, 351–361.
117. Prudêncio, M.; Rohovec, J.; Peters, J. A.; Tocheva, E.; Boulanger, M. J.; Murphy, M. E. P.; Hupkes, H.-J.; Kusters, W.; Impagliazzo, A.; Ubbink, M. *Chem. – Eur. J.* **2004**, *10*, 3252–3260.
118. Consortium, T. U. *Nucleic Acids Res.* **2018**, *47*, D506–D515.
119. Yang, Y.; Wang, J.-T.; Pei, Y.-Y.; Su, X.-C. *Chem. Commun.* **2015**, *51*, 2824–2827.
120. Jiang, W.-X.; Gu, X.-H.; Dong, X.; Tang, C. *J. Biomol. NMR* **2017**, *67*, 273–282.
121. Liu, W.-M.; Keizers, P. H. J.; Hass, M. A. S.; Blok, A.; Timmer, M.; Sarris, A. J. C.; Overhand, M.; Ubbink, M. *J. Am. Chem. Soc.* **2012**, *134*, 17306–17313.
122. Liu, W.-M.; Skinner, S. P.; Timmer, M.; Blok, A.; Hass, M. A. S.; Filippov, D. V.; Overhand, M.; Ubbink, M. *Chem. – Eur. J.* **2014**, *20*, 6256–6258.
123. Lee, M. D.; Dennis, M. L.; Swarbrick, J. D.; Graham, B. *Chem. Commun.* **2016**, *52*, 7954–7957.
124. Lee, M. D.; Dennis, M. L.; Graham, B.; Swarbrick, J. D. *Chem. Commun.* **2017**, *53*, 13205–13208.
125. Voretsky, A.; Gaponenko, V.; Rosevear, P. R. *FEBS Lett.* **2002**, *528*, 189–192.
126. Leonov, A.; Voigt, B.; Rodríguez-Castañeda, F.; Sakhail, P.; Griesinger, C. *Chem. – Eur. J.* **2005**, *11*, 3342–3348.
127. Haber, P.; Rodríguez-Castañeda, F.; Junker, J.; Becker, S.; Leonov, A.; Griesinger, C. *Org. Lett.* **2006**, *8*, 1275–1278.
128. Parker, D.; Dickens, R. S.; Puschmann, H.; Crossland, C.; Howard, J. A. K. *Chem. Rev.* **2002**, *102*, 1977–2010.
129. Chen, J.-L.; Zhao, Y.; Gong, Y.-J.; Pan, B.-B.; Wang, X. Su, X.-C. *J. Biomol. NMR* **2018**, *70*, 77–92.
130. Man, B.; Su, X.-C.; Liang, H.; Simonsen, S.; Huber, T.; Messerle, B. A.; Otting, G. *Chem. – Eur. J.* **2010**, *16*, 3827–3832.
131. Jia, X.; Maleckis, A.; Huber, T.; Otting, G. *Chem. – Eur. J.* **2011**, *17*, 6830–6836.
132. Li, Q.-F.; Yang, Y.; Maleckis, A.; Otting, G.; Su, X.-C. *Chem. Commun.* **2012**, *48*, 2704–2706.
133. Swarbrick, J. D.; Ung, P.; Chhabra, S.; Graham, B. *Angew. Chem. Int. Ed. Engl.* **2011**, *50*, 4403–4406.
134. Swarbrick, J. D.; Ung, P.; Su, X.-C.; Maleckis, A.; Chhabra, S.; Huber, T.; Otting, G.; Graham, B. *Chem. Commun.* **2011**, *47*, 7368–7370.
135. Yagi, H.; Maleckis, A.; Otting, G. *J. Biomol. NMR* **2013**, *55*, 157–166.
136. Loh, C.-T.; Graham, B.; Abdelkader, E. H.; Tuck, K. L.; Otting, G. *Chem. – Eur. J.* **2015**, *21*, 5084–5092.
137. Yang, Y.; Li, Q.-F.; Cao, C.; Huang, F.; Su, X.-C. *Chem. – Eur. J.* **2013**, *19*, 1097–1103.
138. Huang, F.; Pei, Y.-Y.; Zuo, H.-H.; Chen, J.-L.; Yang, Y.; Su, X.-C. *Chem. – Eur. J.* **2013**, *19*, 17141–17149.
139. Tweedle, M. F.; Hagan, J. J.; Kumar, K.; Mantha, S.; Chang, C. A. *Magn. Reson. Imaging* **1991**, *9*, 409–415.
140. Maecke, H. R.; Reubi, J. C. *J. Nucl. Med.* **2011**, *52*, 841–844.
141. Stasiuk, G. J.; Long, N. J. *Chem. Commun.* **2013**, *49*, 2732–2746.
142. Opina, A. C. L.; Strickland, M.; Lee, Y.-S.; Tjandra, N.; Byrd, R. A.; Swenson, R. E.; Vasalatiy, O. *Dalton Trans.* **2016**, *45*, 4673–4687.
143. Opina, A. C. L.; Strickland, M.; Lee, Y. S.; Tjandra, N.; Swenson, R. E.; Vasalatiy, O. *Inorg. Chem.* **2019**, *58*, 15788–15800.
144. Müntener, T.; Thommen, F.; Joss, D.; Kottelat, J.; Prescimone, A.; Häussinger, D. *Chem. Commun.* **2019**, *55*, 4715–4718.
145. Suturina, E. A.; Häussinger, D.; Zimmermann, K.; Garbuio, L.; Yulikov, M.; Jeschke, G.; Kuprov, I. *Chem. Sci.* **2017**, *8*, 2751–2757.
146. Neumann, C. N.; Hooker, J. M.; Ritter, T. *Nature* **2016**, *534*, 369–373.
147. Vlasie, M. D.; Comuzzi, C.; van den Nieuwendijk, A. M. C. H.; Prudêncio, M.; Overhand, M.; Ubbink, M. *Chem. – Eur. J.* **2007**, *13*, 1715–1723.
148. Keizers, P. H. J.; Desreux, J. F.; Overhand, M.; Ubbink, M. *J. Am. Chem. Soc.* **2007**, *129*, 9292–9293.
149. Brath, U.; Swamy, S. I.; Veiga, A. X.; Tung, C.-C.; Van Petegem, F.; Erdélyi, M. *J. Am. Chem. Soc.* **2015**, *137*, 11391–11398.
150. Gao, J.; Liang, E.; Ma, R.; Li, F.; Liu, Y.; Liu, J.; Jiang, L.; Li, C.; Dai, H.; Wu, J.; Su, X.; He, W.; Ruan, K. *Angew. Chem. Int. Ed. Engl.* **2017**, *56*, 12982–12986.
151. Xu, D.; Li, B.; Gao, J.; Liu, Z.; Niu, X.; Nshogozu, G.; Zhang, J.; Wu, J.; Su, X.-C.; He, W.; Ma, R.; Yang, D.; Ruan, K. *J. Phys. Chem. Lett.* **2018**, *9*, 3361–3367.
152. Chen, W.-N.; Nitsche, C.; Pilla, K. B.; Graham, B.; Huber, T.; Klein, C. D.; Otting, G. *J. Am. Chem. Soc.* **2016**, *138*, 4539–4546.
153. Silverman, D. N.; Lindskog, S. *Acc. Chem. Res.* **1988**, *21*, 30–36.
154. Zheng, Y. J.; Merz, K. M. *J. Am. Chem. Soc.* **1992**, *114*, 10498–10507.
155. Supuran, C. T. *Nat. Rev. Drug Discov.* **2008**, *7*, 168–181.
156. Krishnamurthy, V. M.; Kaufman, G. K.; Urbach, A. R.; Gitlin, I.; Gudiksen, K. L.; Weibel, D. B.; Whitesides, G. M. *Chem. Rev.* **2008**, *108*, 946–1051.
157. Eriksson, A. E.; Jones, T. A.; Liljas, A. *Proteins Struct. Funct. Bioinf.* **1988**, *4*, 274–282.
158. Supuran, C. T. *Bioorg. Med. Chem. Lett.* **2010**, *20*, 3467–3474.

159. Schmid, M.; Nogueira, E. S.; Monnard, F. W.; Ward, T. R.; Meuwly, M. *Chem. Sci.* **2012**, *3*, 690–700.
160. Kanamori, K.; Roberts, J. D. *Biochemistry* **1983**, *22*, 2658–2664.
161. Alward, W. L. M. *N. Engl. J. Med.* **1998**, *339*, 1298–1307.
162. Bunzel, H. A.; Garrabou, X.; Pott, M.; Hilvert, D. *Curr. Opin. Struct. Biol.* **2018**, *48*, 149–156.
163. Zeymer, C.; Hilvert, D. *Annu. Rev. Biochem.* **2018**, *87*, 131–157.
164. Arnold, F. H. *Angew. Chem. Int. Ed. Engl.* **2019**, *58*, 14420–14426.
165. Monnard, F. W.; Heinisch, T.; Nogueira, E. S.; Schirmer, T.; Ward, T. R. *Chem. Commun.* **2011**, *47*, 8238–8240.
166. Monnard, F. W.; Nogueira, E. S.; Heinisch, T.; Schirmer, T.; Ward, T. R. *Chem. Sci.* **2013**, *4*, 3269–3274.
167. Heinisch, T.; Pellizzoni, M.; Dürrenberger, M.; Tinberg, C. E.; Köhler, V.; Klehr, J.; Häussinger, D.; Baker, D.; Ward, T. R. *J. Am. Chem. Soc.* **2015**, *137*, 10414–10419.
168. Rebelein, J. G.; Cotellet, Y.; Garabedian, B.; Ward, T. R. *ACS Catal.* **2019**, *9*, 4173–4178.
169. Bereau, T.; Kramer, C.; Monnard, F. W.; Nogueira, E. S.; Ward, T. R.; Meuwly, M. *J. Phys. Chem. B.* **2013**, *117*, 5460–5471.
170. Benej, M.; Pastorekova, S.; Pastorek, J. Carbonic Anhydrase IX: Regulation and Role in Cancer. In *Carbonic Anhydrase: Mechanism, Regulation, Links to Disease, and Industrial Applications*; Frost, S. C., McKenna, R., Eds.; Springer Netherlands: Dordrecht, 2014; pp 199–219.
171. Zhao, J.; Kajetanowicz, A.; Ward, T. R. *Org. Biomol. Chem.* **2015**, *13*, 5652–5655.
172. Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R. G.; Wyckoff, H.; Phillips, D. C. *Nature* **1958**, *181*, 662–666.
173. Merk, A.; Bartesaghi, A.; Banerjee, S.; Falconieri, V.; Rao, P.; Davis, M. I.; Pragani, R.; Boxer, M. B.; Earl, L. A.; Milne, J. L. S.; Subramaniam, S. *Cell* **2016**, *165*, 1698–1707.
174. Wang, J.; Sanchez-Rosello, M.; Acena, J. L.; del Pozo, C.; Sorochinsky, A. E.; Fustero, S.; Soloshonok, V. A.; Liu, H. *Chem. Rev.* **2014**, *114*, 2432–2506.
175. Avvaru, B. S.; Kim, C. U.; Sippel, K. H.; Gruner, S. M.; Agbandje-McKenna, M.; Silverman, D. N.; McKenna, R. *Biochemistry* **2010**, *49*, 249–251.
176. Eddy, M. T.; Didenko, T.; Stevens, R. C.; Wuthrich, K. *Structure* **2016**, *24*, 2190–2197.
177. Kim, C. Y.; Chang, J. S.; Doyon, J. B.; Baird, T. T.; Fierke, C. A.; Jain, A.; Christianson, D. W. *J. Am. Chem. Soc.* **2000**, *122*, 12125–12134.
178. Kurland, R. J.; McGarvey, B. R. *J. Magn. Reson.* **1970**, *2*, 286–301.
179. Oliphant, T. E. *Comput. Sci. Eng.* **2007**, *9*, 10–20.
180. Chandra, P. P.; Jain, A.; Sapse, A. M. *J. Mol. Model.* **2004**, *10*, 1–5.
181. Gómez-Reyes, J. F.; Ariza-Castolo, A. *J. Math. Chem.* **2013**, *51*, 1961–1980.

5 Part III

5.1 Induction of strong paramagnetic effects in RNA molecules by covalent lanthanide tagging for biomolecular NMR

This project was performed in collaboration with Kevin Erharter and Prof. Dr. Christoph Kreutz from the University of Innsbruck.

5.1 Induction of strong paramagnetic effects in RNA molecules by covalent lanthanide tagging for biomolecular NMR

5.1.1 Abstract

Accurate solution structures of biologically relevant RNA molecules are highly desirable for future drug development and basic research. Herein, we report for the first time the use of a rigid, DOTA-based lanthanide chelator that leads upon covalent attachment to an amino methyl modified uridine base to pseudocontact shifts and thus to precise structural restraints on RNA molecules in solution NMR experiments. The incorporated isothiocyanate activator moiety enables convenient ligation of the tag to amino methyl modified uridine nucleotides incorporated within RNA molecules of interest, while the strongly paramagnetic and highly rigidified chelator ensures the generation of large paramagnetic effects in order to obtain precise structural restraints. The successful transfer of the pseudocontact shift methodology from proteins to RNA opens up new avenues for structural characterization of RNA molecules.

5.1.2 Introduction and aim of the study

Precise structural restraints in solution are a key prerequisite for the understanding of the cellular machinery, the underlying processes as well as changes in structure and ligand binding of biomacromolecules.¹⁻² Solution-state nuclear magnetic resonance spectroscopy (NMR) and in particular paramagnetic NMR spectroscopy have shown to be of high value in order to obtain long-range structural restraints for the investigation of the solution structure and dynamics of biomacromolecules.³⁻⁷ More specifically, it was shown that pseudocontact shifts (PCSs) and residual dipolar couplings (RDCs) enable the investigation of protein-ligand⁸ as well as protein-protein complexes,^{7, 9} characterization of interactions between glycans and proteins,¹⁰ mechanistic studies of neurotransmitter release mediated by Synaptotagmin 1-SNARE complexes¹¹ as well as the localization of ligands within a protein scaffold.¹²⁻¹⁵

The successful transfer of the PCS methodology of protein to RNA molecules displays a highly promising and interesting development, since the structural elucidation of RNA molecules, especially the ones that are considered as potential drug targets, will become highly desirable in future drug discovery. The PCS methodology could help to quickly screen drug-RNA interactions and allow to describe the orientation of the drug molecule in the RNA binding pocket, which in turn is a very valuable information to improve the design of drug molecules. However, although the tagging of biomacromolecules and subsequent analysis of the structural restraints is today well established for proteins,^{3-4, 16-21} structural analysis of RNA

targets using paramagnetic restraints induced by direct and covalent tagging with a lanthanide chelating tag (LCT) has, to the best of our knowledge, not been reported yet.

Reported approaches for investigation of protein-DNA,²⁰ protein-DNA mimic²² complexes and monitoring of RNA solution structures²³ rely on paramagnetically tagged reporter proteins. Furthermore, Wu et al. demonstrated for DNA strands that an LCT bearing a bromoacetamide moiety can successfully be incorporated via conjugation to a phosphorothioate²⁴ and Täubert et al. showed that labeling of a modified nucleotide incorporated in DNA is feasible.²⁵ Unfortunately, ligation of a LCT to a chiral phosphorothioate group leads to two diastereomers upon tagging that have to be separated, if possible, by reverse-phase HPLC or anion-exchange chromatography prior to acquisition of NMR experiments.²⁴ An alternative would be to introduce the phosphorothioate at the 5' end of the sequence that would lead to only one single isomer, however, an averaging of the PCSs due to the motions of the LCT can be expected.

Due to the enhanced flexibility and more complex structural variety of RNA molecules when compared to DNA, strongly paramagnetic and covalently and tightly bound LCTs are urgently needed in order to induce large paramagnetic effects and thus precise long-range structural restraints. The specific labeling techniques and RNA modifications developed by the Kreuz research group²⁶⁻²⁹ enable the site-specific ligation of LCTs to stable isotope labeled RNA strands and provide therefore a crucial tool in order to acquire NMR experiments with excellent signal-to-noise ratio and facilitated assignment of the investigated signals. With these prerequisites in hand, we envisaged a highly rigidified, eight-fold, stereospecifically methyl-substituted DOTA-derived LCT bearing an isothiocyanate activator moiety that can be ligated to an amino methyl modified uridine nucleotide in the RNA strand of interest (Ln-DOTA-M8-ITC, Figure 1).

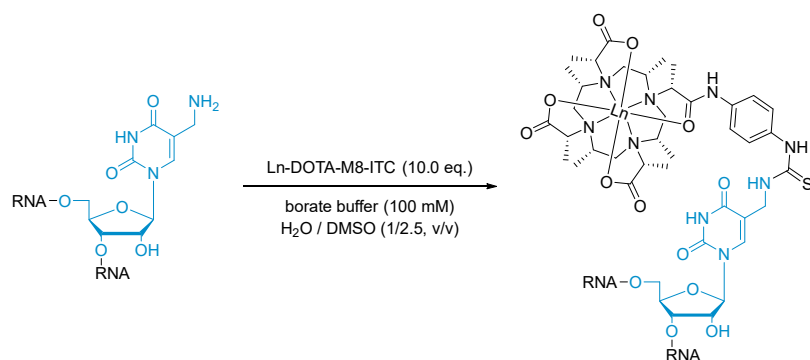


Figure 1: Labeling reaction of an RNA target incorporating an amino methyl modified uridine nucleotide using Ln-DOTA-M8-ITC.

Upon successful conjugation of the LCT to the RNA strand of interest, paramagnetic effects can be obtained via subsequent ¹H-¹³C and ¹H-¹⁵N HSQC experiments. Upon analysis of the

obtained PCS, the suitability of the approach for generation of highly valuable structural restraints from 2D NMR experiments was envisioned to be estimated.

5.1.3 Results and discussion

In order to achieve a successful incorporation of an aminomethyl modified uridine base into the RNA molecule of interest, a synthetic access to 5-aminomethyl uridine phosphoramidite building block **6** was established (Figure 2).

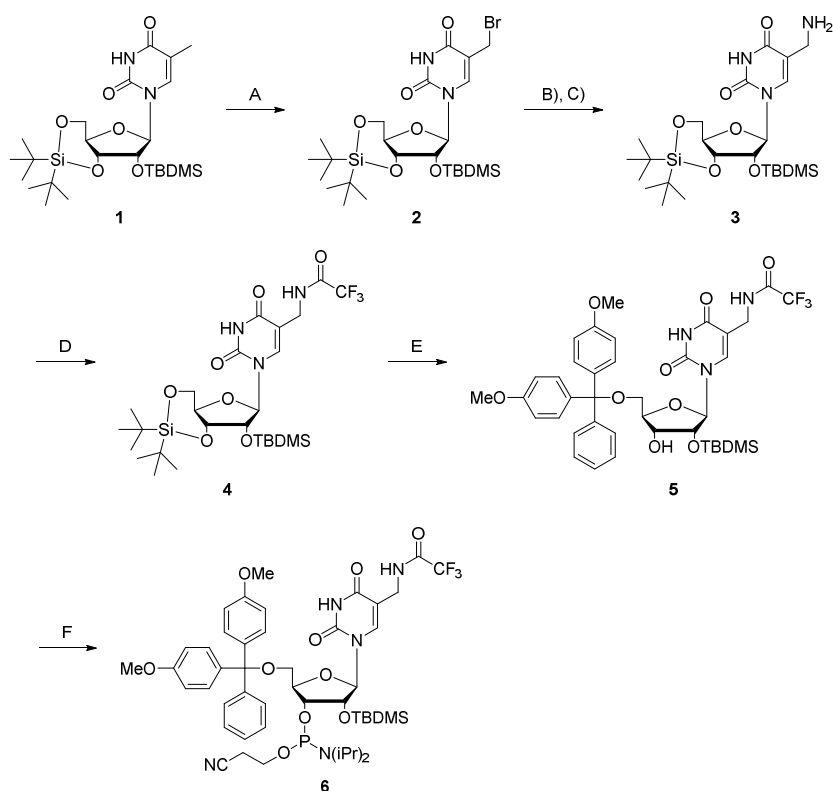


Figure 2: Synthetic access to 5-aminomethyl-modified uridine phosphoramidite **6**. A) *N*-bromosuccinimide (NBS), azobisisobutyronitrile (AIBN) in benzene, 100°C, 4h. B) NaN_3 in DMF, rt, 16 h, 80%. C) PPh_3 , H_2O in benzene, rt, 4h, 77%. D) trifluoroacetic anhydride in pyridine, 0 °C, 30 min, 90%. E) HF-pyridine in CH_2Cl_2 , 0 °C, 2.5 h, then DMT-Cl in pyridine, rt, 12 h, 88%. F) CTIP, BTT in acetonitrile, rt, 12 h, 68%.

The synthesis of the newly developed LCT (Ln-DOTA-M8-ITC) was then started from a precursor molecule with four lactic acid side arms (**7**), three of which were *tert*-butyl protected.³⁰⁻³¹ Installation of the aromatic linker spacer by a HATU-mediated coupling (**8**) was followed by deprotection of the *tert*-butyl protecting groups in order to yield the free ligand **9** (Figure 3).

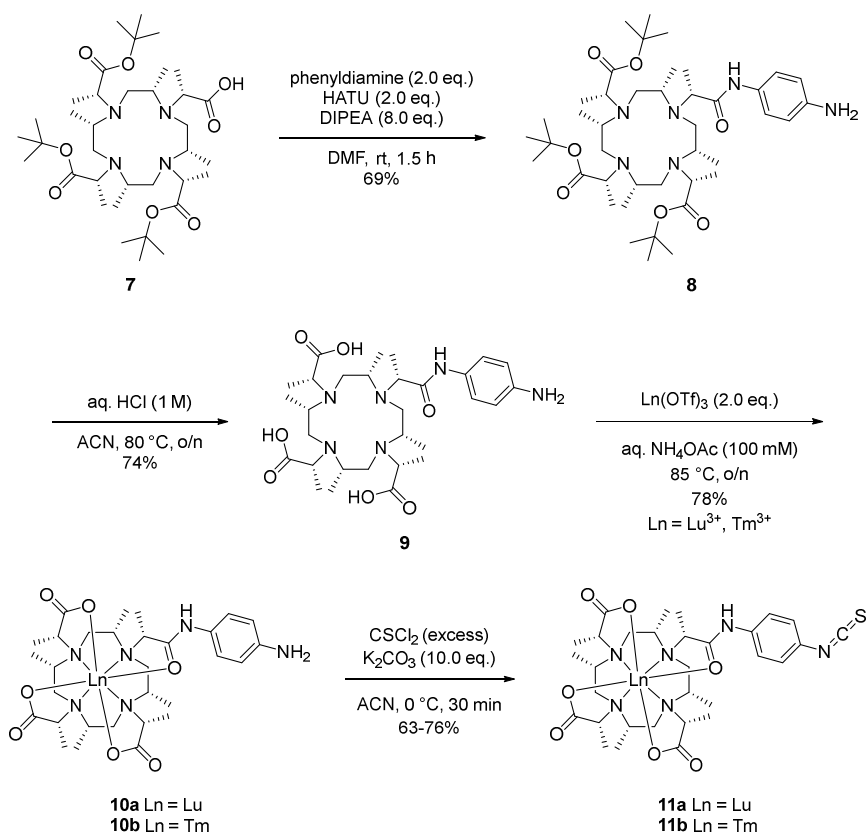


Figure 3: Synthetic route to the newly developed Ln-DOTA-M8-ITC LCT bearing an isothiocyanate activator moiety for tagging of amino groups incorporated in RNA bases.

Upon complexation of a suitable lanthanide ion (Tm^{3+} : paramagnetic, Lu^{3+} : diamagnetic reference) to yield lanthanide complex **10**, the isothiocyanate activator moiety was installed in the last step of the synthetic route due to its potential instability towards hydrolysis in the metalation step under heating conditions. The synthetic route to the newly developed Ln-DOTA-M8-ITC LCT (**11**) provides a convenient access in an overall yield of 30% starting from the threefold *tert*-butyl substituted, protected precursor **7**.

The amino functionalized uridine base was introduced into the 36-nt stem loop (SL) C^A from the Moloney murine leukemia virus 5'-leader, which was earlier addressed by Tjandra and co-workers via the U1A LCT approach.²³ More specifically, $^{13}\text{C}/^{15}\text{N}$ labels were introduced as well as a modified nucleotide, i.e. 5-aminomethyl uridine, at position 16 (Figure 4A). The modified nucleotide was inserted at position 16 due to the optimal positioning of the base in a loop structure that gives a readily accessible tagging site for conjugation of the Ln-DOTA-M8-ITC LCT to the SLC^A RNA molecule.

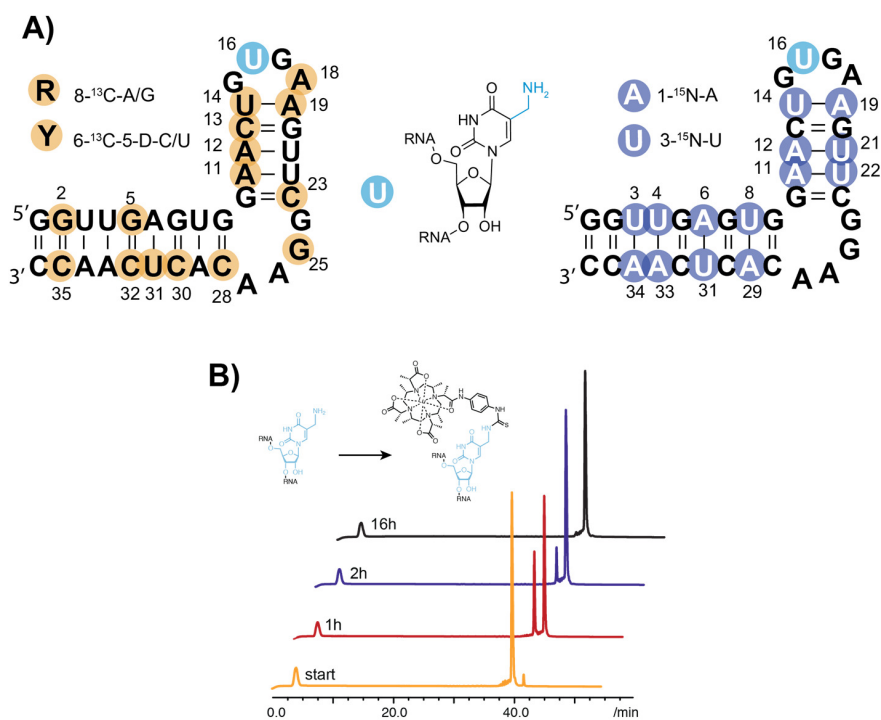


Figure 4: A) 36-nt SLC^A with ¹³C (orange) and ¹⁵N (dark blue) labels and the 5-aminomethyl uridine residue at position 16 (light blue). B) Tagging reaction monitored by anion exchange chromatography (10 eq. Lu-DOTA-M8-ITC in 100 mM borate buffer with pH 8.6 in H₂O / DMSO (1/2.5, v/v)). Figure adapted from the original Figure by Christoph Kreutz (University of Innsbruck).

Using the synthesized Ln-DOTA-M8-ITC LCT complexed either with Tm³⁺ or Lu³⁺, the primary amino group of U16 was used as an anchor point to attach the LCT via a thioureido functionality. The successful conjugation reaction is exemplarily shown for the lutetium variant (Figure 4B). The ligation of Ln-DOTA-M8-ITC to the RNA strand was complete after 5 h. After the successful ligation reaction, we determined 36 ¹H, ¹³C and ¹⁵N PCSs for the SLC^A construct, which were then used to obtain the anisotropy parameters and position of the metal center (Figure 5, Table 1). A comparison of NMR experiments of RNA labeled with Lu-DOTA-M8-ITC and without conjugated LCT shows that there are no global structural distortions of the RNA arising due to conjugation with Ln-DOTA-M8-ITC (Figure S12).

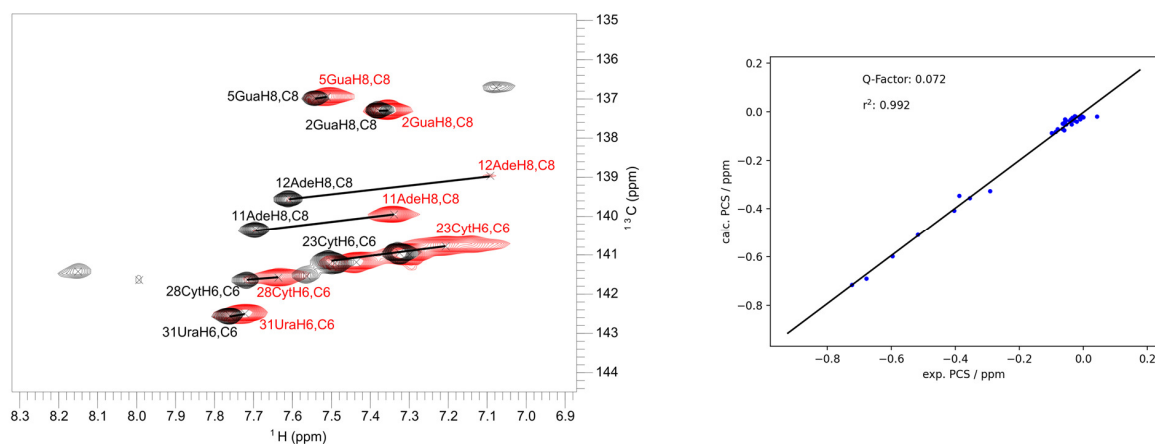


Figure 5: Overlay of ^1H - ^{13}C TROSY HSQC experiments with selected PCS on the SLC^{A} (left) and correlation plot of experimental and back-calculated PCS (right).

Table 1: Induced axial and rhombic components of the paramagnetic susceptibility tensors ($\Delta\chi_{\text{ax}}$ and $\Delta\chi_{\text{rh}}$, in 10^{-32} m^3), metal position in PDB coordinate frame (X_{metal} , Y_{metal} , Z_{metal} , in \AA), Euler angles (α , β , γ , in $^\circ$) and quality factor (Q , mathematical definition in SI, p. 1) on a Tm-DOTA-M8-ITC- SLC^{A} construct (PDB in Marchant et al.³²) at pH 6.5 and 298 K (Lu-DOTA-M8-ITC- SLC^{A} construct used as reference).

N ^o PCS	$\Delta\chi_{\text{ax}}$ (10^{-32} m^3)	$\Delta\chi_{\text{rh}}$ (10^{-32} m^3)	X_{metal} (\AA)	Y_{metal} (\AA)	Z_{metal} (\AA)	α ($^\circ$)	β ($^\circ$)	γ ($^\circ$)	Q (%)
36	21.5	8.7	23.9	-4.8	41.9	38.6	91.0	30.5	7.2

The experiments yielded anisotropy parameters of $\Delta\chi_{\text{ax}} = 21.5 \times 10^{-32} \text{ m}^3$ and $\Delta\chi_{\text{rh}} = 8.7 \times 10^{-32} \text{ m}^3$ with a quality factor $Q = 0.072$ (Table 1), which is well comparable both in the axial magnitude and quality of the fit to the results obtained with Ln-DOTA-M8-(4R4S)-SSPy LCT on ubiquitin S57C constructs in protein NMR studies ($\Delta\chi_{\text{ax}} = 19.6 \times 10^{-32} \text{ m}^3$, $\Delta\chi_{\text{rh}} = 3.0 \times 10^{-32} \text{ m}^3$, $Q = 0.062$, 68 PCSs).³¹

The distance of the fitted metal center position to the aminomethyl modification amounts to 8.2 \AA , which is a realistic value for the used thioureido linkage including the aromatic spacer. Absolute deviation values of axial and rhombic components of PCS-derived tensors were accounted for by a Monte-Carlo simulation with a perturbation level of 0.1 \AA on the RNA structure and 10000 iterations. This analysis yielded anisotropy parameters of $\Delta\chi_{\text{ax}} = 21.5 \pm 3.5 \times 10^{-32} \text{ m}^3$ and $\Delta\chi_{\text{rh}} = 8.8 \pm 2.1 \times 10^{-32} \text{ m}^3$ for the investigated Tm-DOTA-M8-ITC- SLC^{A} construct.

The results with the SLC^{A} RNA strand, i.e. the successful synthesis and ligation reaction of Ln-DOTA-M8-ITC as well as the derivation of the anisotropy parameters from the experimental constraints by a successful fit of the experimental constraints, are highly promising. The introduction of methyl-substituted, DOTA-based LCTs for tagging of RNA constructs paves the

avenue for the use of anisotropic information from RNA NMR experiments in solution state in the investigation of structure and dynamics of suitable RNA targets.

One drawback of our current approach, however, is that the maximal PCS reaches only 0.72 ppm. We attribute this to the residual rotational and translational degrees of freedom of the linker. In order to reach an even higher precision and the possibility to characterize larger RNA targets, a significantly shortened and rigidified linker could be used. The expected rigidification would lead to a decreased and thus more favorable rotational averaging of the measured paramagnetic effects and therefore to both a higher precision and a larger distance range that can be monitored using the paramagnetic LCTs.^{17, 33-34}

In order to obtain more insights into the motional averaging of the Ln-DOTA-M8-ITC LCT when attached to the RNA strand, MD simulations were scheduled. An initial estimate of the non-averaged, intrinsic anisotropy parameters on the free tag was obtained from 1-dimensional ¹H NMR experiments in combination with DFT structures of Tm- and Dy-DOTA-M8-ITC. Thereby, values of $\Delta\chi_{ax} = 95.2 \times 10^{-32} \text{ m}^3$ and $\Delta\chi_{rh} = 2.6 \times 10^{-32} \text{ m}^3$ for the Tm- ($Q = 0.024$, 12 PCSs) and $\Delta\chi_{ax} = -69.8 \times 10^{-32} \text{ m}^3$ and $\Delta\chi_{rh} = -1.0 \times 10^{-32} \text{ m}^3$ for the Dy-loaded chelator ($Q = 0.020$, 12 PCSs) were obtained that can serve as starting point for future MD simulations (Table S2-S3, Figure S15-S18). The simulations are then expected to yield insights into the dynamics of the RNA by taking into account the intrinsic anisotropy parameters of the LCT, the effective anisotropy parameters obtained on the RNA-LCT conjugate and motional data from MD simulations.

5.1.4 Conclusions and Outlook

To conclude, for the first time direct and covalent tagging of RNA molecules with a strongly paramagnetic lanthanide complex that allows for generation of sizeable PCSs and hence precise structural restraints is reported. The isothiocyanate activator moiety enables convenient and fast tagging of amino methyl modified uridine nucleotides. In combination with the highly rigidified DOTA-chelator that leads to sizeable PCSs, the Ln-DOTA-M8-ITC LCT enables future applications for the investigation of the solution structure of RNA molecules and their RNA-RNA and RNA-ligand complexes. The modified uridine nucleotide can be incorporated in any RNA strand suitable for solid-phase synthesis and thus, the presented approach can be considered as a general approach towards the structural elucidation of RNA targets in solution using paramagnetic NMR spectroscopy. In addition to Lu- and Tm-DOTA-M8-ITC, the Dy-complex was recently synthesized and is currently under investigation for both the application on the 36-nt SLC^A RNA strand and a larger 64-nt RNA molecule incorporating a ligand. Furthermore, MD simulations are on the way for both RNA constructs in order to investigate the individual contributions of the non-averaged tensors to the average tensor obtained by solution NMR. In future, the synthesis of novel, very short pyridine activator

moieties could provide larger PCSs and thereby enlarge the obtained anisotropy parameters. Due to the high interest in structures of RNA as well as RNA-ligand complexes in future drug discovery, the presented method holds great promise for the structural characterization and investigation of ligand-binding properties of RNA molecules under physiological conditions.

5.1.5 Supporting information

General Remarks

Unless otherwise stated, reactions were performed under argon atmosphere and chemicals were used as received without further purification. All reactions were executed in anhydrous solvents.

NMR experiments were performed at a temperature of 298 K on a Bruker Avance III HD NMR spectrometer operating at 600 MHz equipped with a cryogenic QCI-F probehead. ESI-MS spectra were recorded on a Shimadzu LCMS-2020 liquid chromatograph mass spectrometer. HRMS spectra were measured on a Bruker MaXis 4G HiRes ESI Mass Spectrometer.

IR spectra were acquired using a Shimadzu IRTracer-100 equipped with a QATR 10 single reflection ATR accessory. The obtained spectra were analyzed with the Shimadzu LabSolutions IR software package.

(*R*)-2-((2*S*,5*S*,8*S*,11*S*)-4,7,10-tris((*R*)-1-(*tert*-butoxy)-1-oxopropan-2-yl)-2,5,8,11-tetramethyl-1,4,7,10-tetraazacyclododecan-1-yl)propanoic acid was synthesized according to the published procedures by Joss et al.³¹ ¹H-¹⁵N SOFAST-HMQC and ¹H-¹³C BEST-TROSY-HSQC experiments were measured in 15 mM phosphate buffer (25 mM NaCl, 10% D₂O) with pH 6.5 at a temperature of 298 K on a 600 MHz Bruker Avance II+ NMR spectrometer equipped with a liquid N₂ cooled TCI Prodigy probe. The obtained NMR spectra were assigned using CcpNmr Analysis.³⁵ The tensor properties were then obtained by fitting the signals to the ¹H-¹³C RDC optimized structure of SLC^A³² using Numbat³⁶. Absolute deviation values of axial and rhombic components of PCS-derived tensors were estimated by a Monte-Carlo simulation with a perturbation level of 0.1 Å on the RNA structure and 10000 iterations. The metal center was found in a distance of 8.2 Å from the aminomethyl group C of the modified uridine base (16). Q-factors were calculated using the following equation:

$$Q = \frac{\sqrt{\sum (PCS_{exp} - PCS_{calc})^2}}{\sqrt{\sum (PCS_{exp})^2}}$$

HPLC-MS conditions (synthesis)

Analytical HPLC measurements were performed on a Shimadzu LC system (LC-20AT prominence liquid chromatograph, SIL-20A HT prominence auto sampler, CTO-20AC prominence column oven, CBM-20A communications bus module, SPD-20A prominence UV/VIS detector ($\lambda = 254$ and 280 nm)) combined with a Shimadzu LCMS-2020 liquid chromatograph mass spectrometer. As column for analytical HPLC measurements, a ReproSil-Pur ODS-3, $3 \mu\text{m}$, 150×3 mm, provided by Maisch GmbH was used. Commercial HPLC grade solvents were used and a binary gradient was applied.

Solvent A: Milli-Q water + 0.1% TFA

Solvent B: 90% acetonitrile + 10% Milli-Q water + 0.085% TFA.

HPLC gradient: 95% A (min 0-2), linear gradient 95% A to 100% B (min 2-6), 100% B (min 6-14), linear gradient 100% B to 95% A (min 14-15), 95% A (min 15-22).

Semi-preparative HPLC purification was performed on a Shimadzu LC system (LC-20AT prominence liquid chromatograph, SIL-20A HT prominence auto sampler, CTO-20AC prominence column oven, CBM-20A communications bus module, SPD-20A prominence UV/VIS detector ($\lambda = 254$ and 280 nm), LC-20AD fraction collector) combined with a Shimadzu LCMS-2020 liquid chromatograph mass spectrometer. As column for preparative HPLC purification, a ReproSil-Pur 120 ODS-3, $5 \mu\text{m}$, 150×20 mm, provided by Maisch GmbH was used. Commercial HPLC grade solvents were used and a binary gradient was applied during purification.

Solvent A: Milli-Q water + 0.1% TFA

Solvent B: 90% acetonitrile + 10% Milli-Q water + 0.085% TFA.

HPLC gradient: 95% A (min 0-2), linear gradient 95% A to 100% B (min 2-15), 100% B (min 15-22), linear gradient 100% B to 95% A (min 22-23), 95% A (min 23-25).

Mass spectrometry (RNA)

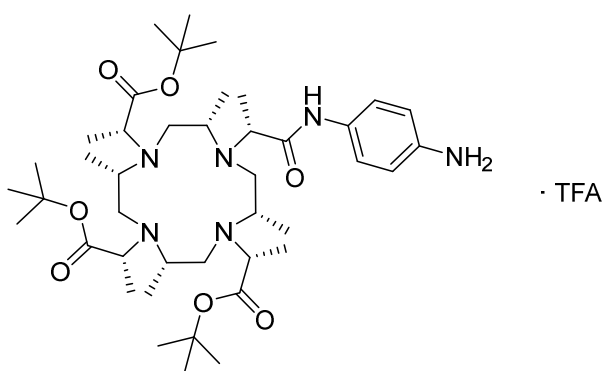
LC-ESI MS analysis was performed using a Finnigan LCQ Advantage MAX ion trap and a ThermoFisher Ultimate 3000 system (ion-pair reversed phase liquid chromatography), equipped with RS pump, RS autosampler, diode array detector and an XBridge C18 column.

Solvent A: 8.6 mM TEA, 100 mM HFIP in Milli-Q water

Solvent B: MeOH

Tri-*tert*-butyl 2,2',2''-((2*S*,5*S*,8*S*,11*S*)-10-((*R*)-1-((4-aminophenyl)amino)-1-oxopropan-2-yl)-2,5,8,11-tetramethyl-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)(2*R*,2'*R*,2''*R*)-tripropionate

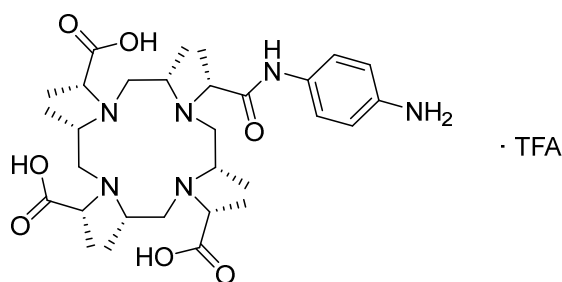
A solution of (*R*)-2-((2*S*,5*S*,8*S*,11*S*)-4,7,10-tris((*R*)-1-(*tert*-butoxy)-1-oxopropan-2-yl)-2,5,8,11-tetramethyl-1,4,7,10-tetraaza-cyclododecan-1-yl)propanoic acid (40.0 mg, 0.058 mmol, 1.0 eq.), DIPEA (0.082 mL, 0.467 mmol, 8.0 eq.) and HATU (44.4 mg, 0.117 mmol, 2.0 eq.) in DMF (2.0 mL) was slowly added to a solution of benzene-1,4-diamine (12.6 mg, 0.117 mmol, 2.0 eq.) in DMF (2.0 mL) and the mixture was stirred overnight. The crude product was purified by prep. HPLC to yield tri-*tert*-butyl 2,2',2''-((2*S*,5*S*,8*S*,11*S*)-10-((*R*)-1-((4-aminophenyl)amino)-1-oxopropan-2-yl)-2,5,8,11-tetramethyl-1,4,7,10-tetraazacyclo-dodecane-1,4,7-triyl)(2*R*,2'*R*,2''*R*)-tripropionate (31.0 mg, 0.040 mmol, 69%) as a white powder.



HRMS: $[M+H]^+$ $C_{42}H_{75}N_6O_7$, m/z (calc.) = 775.5692, m/z (meas.) = 775.5698.

(2*R*,2'*R*,2''*R*)-2,2',2''-((2*S*,5*S*,8*S*,11*S*)-10-((*R*)-1-((4-aminophenyl)amino)-1-oxopropan-2-yl)-2,5,8,11-tetramethyl-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)tripropionic acid

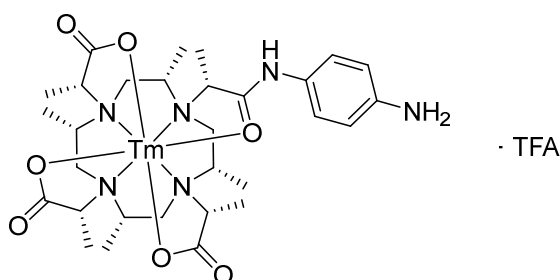
A solution of tri-*tert*-butyl 2,2',2''-((2*S*,5*S*,8*S*,11*S*)-10-((*R*)-1-((4-aminophenyl)amino)-1-oxopropan-2-yl)-2,5,8,11-tetramethyl-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)(2*R*,2'*R*,2''*R*)-tripropionate (24.0 mg, 0.031 mmol, 1.0 eq.) in aq. HCl (1.3 mL, 1 M) and acetonitrile (0.7 mL) was heated to 80 °C overnight. The crude product was purified by prep. HPLC to yield (2*R*,2'*R*,2''*R*)-2,2',2''-((2*S*,5*S*,8*S*,11*S*)-10-((*R*)-1-((4-aminophenyl)amino)-1-oxopropan-2-yl)-2,5,8,11-tetramethyl-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)tripropionic acid (16.0 mg, 0.026 mmol, 85%) as a colorless, glassy solid.



HRMS: [M+H]⁺ C₃₀H₅₁N₆O₇, m/z (calc.) = 607.3814, m/z (meas.) = 607.3805.

Tm-(2*R*,2'*R*,2''*R*)-2,2',2''-((2*S*,5*S*,8*S*,11*S*)-10-((*R*)-1-((4-aminophenyl)amino)-1-oxopropan-2-yl)-2,5,8,11-tetramethyl-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)tripropionate

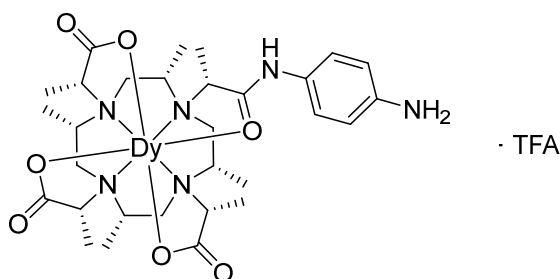
Thulium(III) trifluoromethanesulfonate (6.1 mg, 9.9 μmol , 2.0 eq.) was added to a solution of (2*R*,2'*R*,2''*R*)-2,2',2''-((2*S*,5*S*,8*S*,11*S*)-10-((*R*)-1-((4-aminophenyl)amino)-1-oxopropan-2-yl)-2,5,8,11-tetramethyl-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)tripropionic acid (3.0 mg, 4.9 μmol , 1.0 eq.) in aq. ammonium acetate (2.0 mL, 100 mM) and heated to 85 °C overnight. The crude product was purified by prep. HPLC to yield Tm-(2*R*,2'*R*,2''*R*)-2,2',2''-((2*S*,5*S*,8*S*,11*S*)-10-((*R*)-1-((4-aminophenyl)amino)-1-oxopropan-2-yl)-2,5,8,11-tetramethyl-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)tripropionate (3.0 mg, 3.9 μmol , 78%) as a white powder.



HRMS: $[\text{M}+\text{H}]^+$ $\text{C}_{30}\text{H}_{48}\text{N}_6\text{O}_7\text{Tm}$, m/z (calc.) = 773.2921, m/z (meas.) = 773.2922.

Dy-(2*R*,2'*R*,2''*R*)-2,2',2''-((2*S*,5*S*,8*S*,11*S*)-10-((*R*)-1-((4-aminophenyl)amino)-1-oxopropan-2-yl)-2,5,8,11-tetramethyl-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)tripropionate

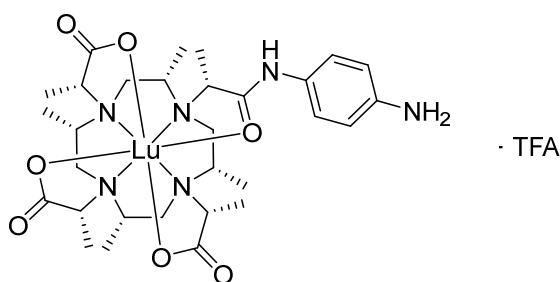
Dysprosium(III) trifluoromethanesulfonate (14.0 mg, 0.023 mmol, 2.0 eq.) was added to a solution of (2*R*,2'*R*,2''*R*)-2,2',2''-((2*S*,5*S*,8*S*,11*S*)-10-((*R*)-1-((4-aminophenyl)amino)-1-oxopropan-2-yl)-2,5,8,11-tetramethyl-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)tripropionic acid (7.0 mg, 0.012 mmol, 1.0 eq.) in aq. ammonium acetate (4.0 mL, 100 mM) and heated to 85 °C overnight. The crude product was purified by prep. HPLC to yield Dy-(2*R*,2'*R*,2''*R*)-2,2',2''-((2*S*,5*S*,8*S*,11*S*)-10-((*R*)-1-((4-aminophenyl)amino)-1-oxopropan-2-yl)-2,5,8,11-tetramethyl-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)tripropionate (5.0 mg, 6.5 μmol , 57%) as a white powder.



HRMS: $[\text{M}+\text{H}]^+$ $\text{C}_{30}\text{H}_{48}\text{DyN}_6\text{O}_7$, m/z (calc.) = 768.2878, m/z (meas.) = 768.2878.

Lu-(2*R*,2'*R*,2''*R*)-2,2',2''-((2*S*,5*S*,8*S*,11*S*)-10-((*R*)-1-((4-aminophenyl)amino)-1-oxopropan-2-yl)-2,5,8,11-tetramethyl-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)tripropionate

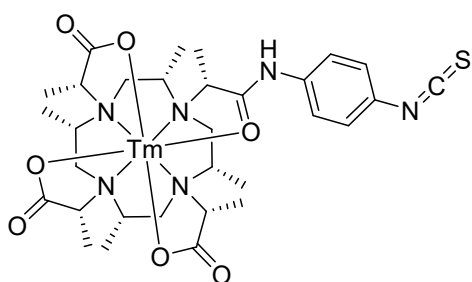
Lutetium(III) trifluoromethanesulfonate (16.4 mg, 0.026 mmol, 2.0 eq.) was added to a solution of (2*R*,2'*R*,2''*R*)-2,2',2''-((2*S*,5*S*,8*S*,11*S*)-10-((*R*)-1-((4-aminophenyl)amino)-1-oxopropan-2-yl)-2,5,8,11-tetramethyl-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)tripropionic acid (8.0 mg, 0.013 mmol, 1.0 eq.) in aq. ammonium acetate (2.0 mL, 100 mM) and heated to 85 °C overnight. The crude product was purified by prep. HPLC to yield Lu-(2*R*,2'*R*,2''*R*)-2,2',2''-((2*S*,5*S*,8*S*,11*S*)-10-((*R*)-1-((4-aminophenyl)amino)-1-oxopropan-2-yl)-2,5,8,11-tetramethyl-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)tripropionate (9.0 mg, 0.012 mmol, 87%) as a white powder.



HRMS: [M+H]⁺ C₃₀H₄₈LuN₆O₇, m/z (calc.) = 779.2987, m/z (meas.) = 779.2973.

Tm-(2*R*,2'*R*,2''*R*)-2,2',2''-((2*S*,5*S*,8*S*,11*S*)-10-((*R*)-1-(4-isothiocyanatophenyl)amino)-1-oxopropan-2-yl)-2,5,8,11-tetramethyl-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)tripropanoate, Tm-DOTA-M8-ITC

Thiophosgene (0.030 mL, 0.388 mmol, 50.0 eq.) was added to a solution of Tm-(2*R*,2'*R*,2''*R*)-2,2',2''-((2*S*,5*S*,8*S*,11*S*)-10-((*R*)-1-(4-aminophenyl)amino)-1-oxopropan-2-yl)-2,5,8,11-tetramethyl-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)tripropanoate (6.0 mg, 7.8 μmol, 1.0 eq.) and potassium carbonate (10.7 mg, 0.078 mmol, 10.0 eq.) in acetonitrile (1.5 mL) at 0 °C and the mixture was stirred for 1 h at 0 °C. The crude product was purified by prep. HPLC to yield Tm-(2*R*,2'*R*,2''*R*)-2,2',2''-((2*S*,5*S*,8*S*,11*S*)-10-((*R*)-1-(4-isothiocyanatophenyl)amino)-1-oxopropan-2-yl)-2,5,8,11-tetramethyl-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)tripropanoate (6.0 mg, 7.3 μmol, 95%) as a white powder.

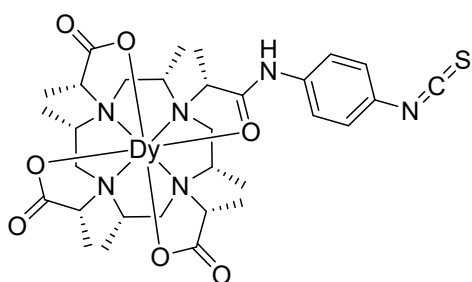


HRMS: [M+H]⁺ C₃₁H₄₆N₆O₇STm, m/z (calc.) = 815.2485, m/z (meas.) = 815.2473.

IR (ν/cm⁻¹): 3397 (ν_{N-H}), 2982 (ν_{C-H}), 2949 (ν_{C-H}), 2093 (ν_{N=C=S}), 1613 (ν_{C=O}), 1597 (ν_{C=C}), 1393 (δ_{C-H}), 839 (δ_{C-H}).

Dy-(2*R*,2'*R*,2''*R*)-2,2',2''-((2*S*,5*S*,8*S*,11*S*)-10-((*R*)-1-((4-isothiocyanatophenyl)amino)-1-oxopropan-2-yl)-2,5,8,11-tetramethyl-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)tripropanoate, Dy-DOTA-M8-ITC

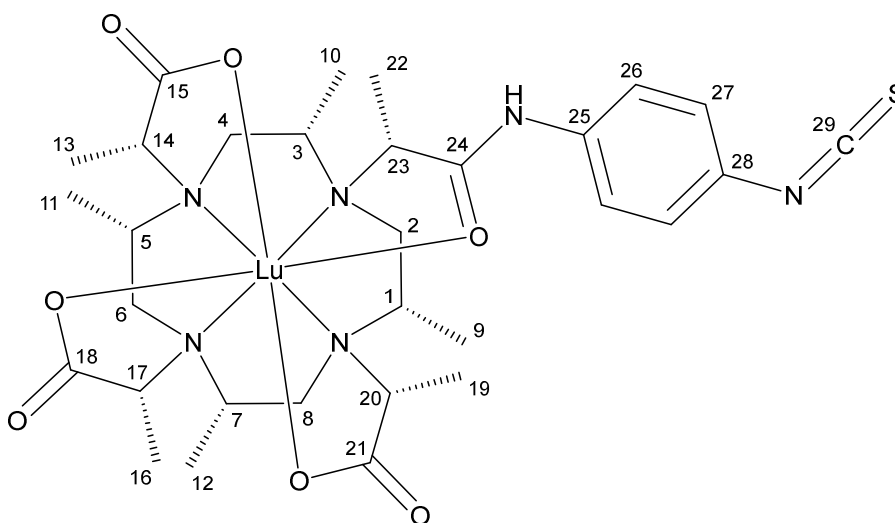
Thiophosgene (0.025 mL, 0.326 mmol, 50.0 eq.) was added to a solution of Dy-(2*R*,2'*R*,2''*R*)-2,2',2''-((2*S*,5*S*,8*S*,11*S*)-10-((*R*)-1-((4-aminophenyl)amino)-1-oxopropan-2-yl)-2,5,8,11-tetramethyl-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)tripropanoate (5.0 mg, 6.5 μ mol, 1.0 eq.) and potassium carbonate (9.0 mg, 0.065 mmol, 10.0 eq.) in acetonitrile (1.5 mL) at 0 °C and the mixture was stirred for 1 h at 0 °C. The crude product was purified by prep. HPLC to yield Dy-(2*R*,2'*R*,2''*R*)-2,2',2''-((2*S*,5*S*,8*S*,11*S*)-10-((*R*)-1-((4-isothiocyanatophenyl)amino)-1-oxopropan-2-yl)-2,5,8,11-tetramethyl-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)tripropanoate (4.0 mg, 4.9 μ mol, 76%) as a white powder.



HRMS: $[M+H]^+$ $C_{31}H_{46}DyN_6O_7S$, m/z (calc.) = 810.2440, m/z (meas.) = 810.2432.

Lu-(2*R*,2'*R*,2''*R*)-2,2',2''-((2*S*,5*S*,8*S*,11*S*)-10-((*R*)-1-((4-isothiocyanatophenyl)amino)-1-oxopropan-2-yl)-2,5,8,11-tetramethyl-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)tripropanoate, Lu-DOTA-M8-ITC

Thiophosgene (0.024 mL, 0.320 mmol, 50.0 eq.) was added to a solution of Lu-(2*R*,2'*R*,2''*R*)-2,2',2''-((2*S*,5*S*,8*S*,11*S*)-10-((*R*)-1-((4-aminophenyl)amino)-1-oxopropan-2-yl)-2,5,8,11-tetramethyl-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)tripropanoate (5.0 mg, 6.4 μ mol, 1.0 eq.) and potassium carbonate (8.9 mg, 0.064 mmol, 10.0 eq.) in acetonitrile (1.5 mL) at 0 °C and the mixture was stirred for 1 h at 0 °C. The crude product was purified by prep. HPLC to yield Lu-(2*R*,2'*R*,2''*R*)-2,2',2''-((2*S*,5*S*,8*S*,11*S*)-10-((*R*)-1-((4-isothiocyanatophenyl)amino)-1-oxopropan-2-yl)-2,5,8,11-tetramethyl-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)tripropanoate (5.0 mg, 6.1 μ mol, 95%) as a white powder.



¹H-NMR (600 MHz, D₂O, δ /ppm): 7.50-7.47 (m, 2H, **H₂₆**), 7.33-7.30 (m, 2H, **H₂₇**), 4.21 (q, ³*J*_{H23-H22} = 7.3 Hz, 1H, **H₂₃**), 3.82-3.73 (m, 3H, **H₁₄**, **H₁₇**, **H₂₀**), 3.19-3.00 (m, 8H, **H_{1ax}**, **H_{2ax}**, **H_{3ax}**, **H_{4ax}**, **H_{5ax}**, **H_{6ax}**, **H_{7ax}**, **H_{8ax}**), 2.74-2.64 (m, 4H, **H_{2eq}**, **H_{4eq}**, **H_{6eq}**, **H_{8eq}**), 1.67 (d, ³*J*_{H22-H23} = 7.4 Hz, 3H, **H₂₂**), 1.44 (d, ³*J*_{H16-H17} = 7.3 Hz, 3H, **H₁₆**), 1.43 (d, ³*J*_{H19-H20} = 7.2 Hz, 3H, **H₁₉**), 1.41 (d, ³*J*_{H13-H14} = 7.2 Hz, 3H, **H₁₃**), 1.18 (d, ³*J*_{H10-H3} = 5.9 Hz, 3H, **H₁₀**), 1.17-1.14 (m, 9H, **H₉**, **H₁₁**, **H₁₂**).

¹³C-NMR (151 MHz, D₂O, δ /ppm, extracted from HSQC and HMBC): 183.3 (**C₁₈**), 183.2 (**C₂₁**), 182.7 (**C₁₅**), 179.5 (**C₂₄**), 134.4 (**C₂₅**), 134.2 (**C₂₉**), 128.5 (**C₂₈**), 126.3 (**C₂₇**), 123.2 (**C₂₆**), 67.4 (**C₁₄**), 67.4 (**C₁₇**), 67.2 (**C₂₀**), 65.7 (**C₂₃**), 61.3 (**C₇**), 61.2 (**C₁**), 61.1 (**C₃**), 61.0 (**C₅**), 45.8 (**C₂**), 45.8 (**C₄**), 45.8 (**C₆**), 45.8 (**C₈**), 14.0 (**C₂₂**), 13.6 (**C₁₆**), 13.6 (**C₁₉**), 13.2 (**C₁₃**), 12.7 (**C₁₀**), 12.6 (**C₉**), 12.6 (**C₁₁**), 12.6 (**C₁₂**).

HRMS: [M+H]⁺ C₃₁H₄₆LuN₆O₇S, m/z (calc.) = 821.2551, m/z (meas.) = 821.2567.

NMR spectra of synthesized compounds

^1H NMR of Lu-DOTA-M8-ITC (D_2O , 298 K, QCI-F, 600 MHz)

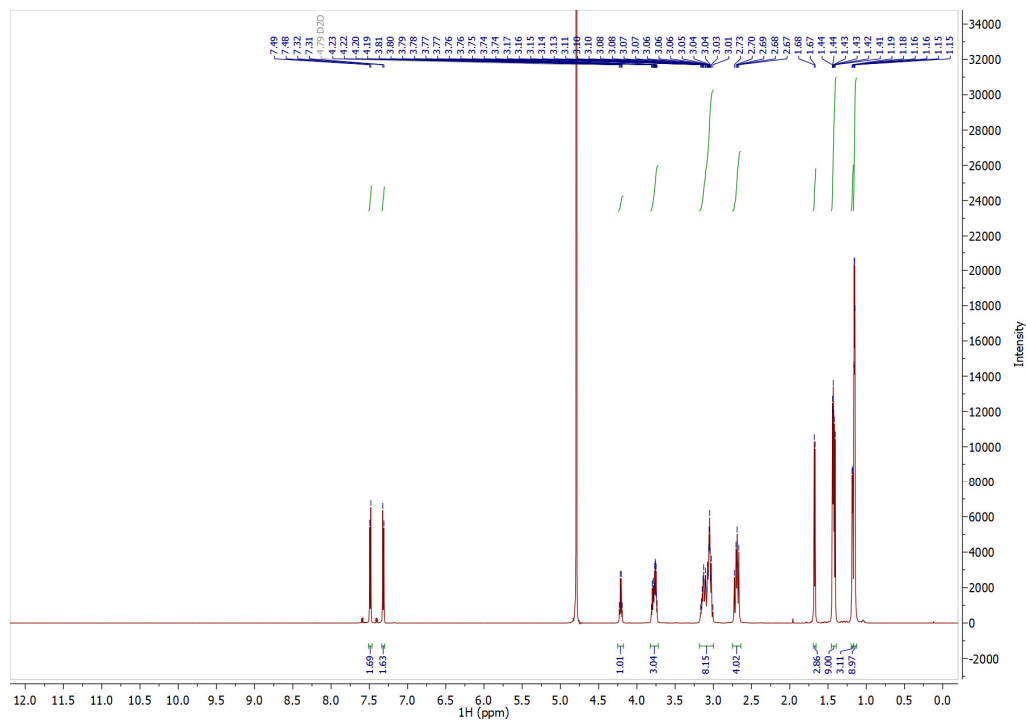


Figure S1: ^1H NMR of Lu-DOTA-M8-ITC (D_2O , 298 K, QCI-F, 600 MHz).

^1H -NMR of Tm-DOTA-M8-ITC (D_2O , 298 K, QCI-F, 600 MHz)

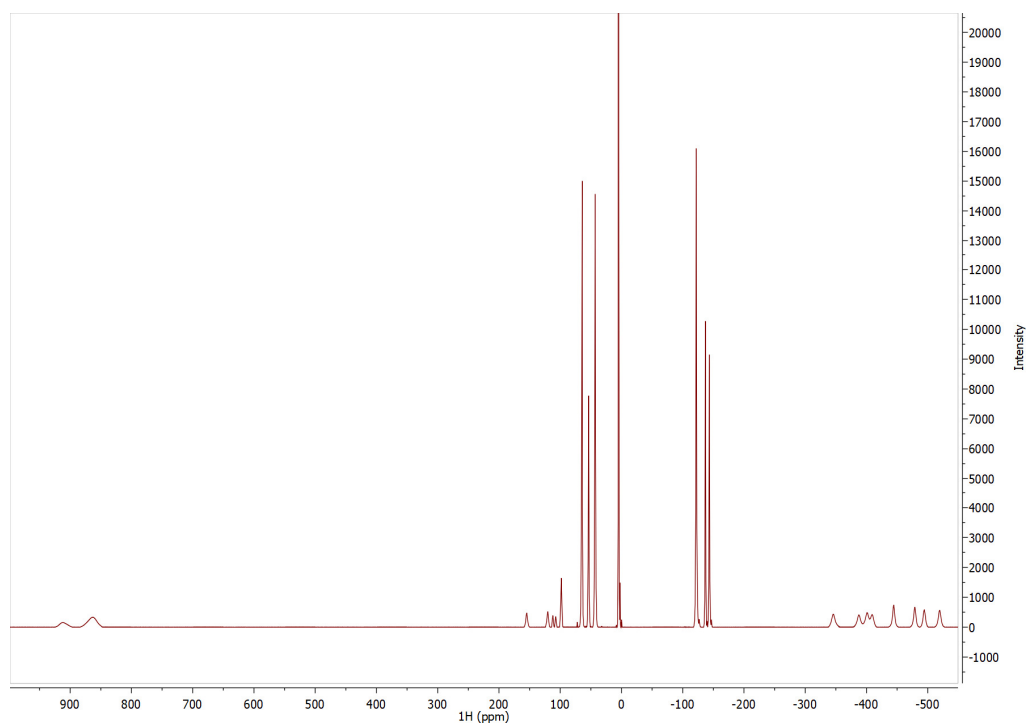


Figure S2: ^1H NMR of Tm-DOTA-M8-ITC (D_2O , 298 K, QCI-F, 600 MHz).

¹H-NMR of Dy-DOTA-M8-ITC (D₂O, 298 K, QCI-F, 600 MHz)

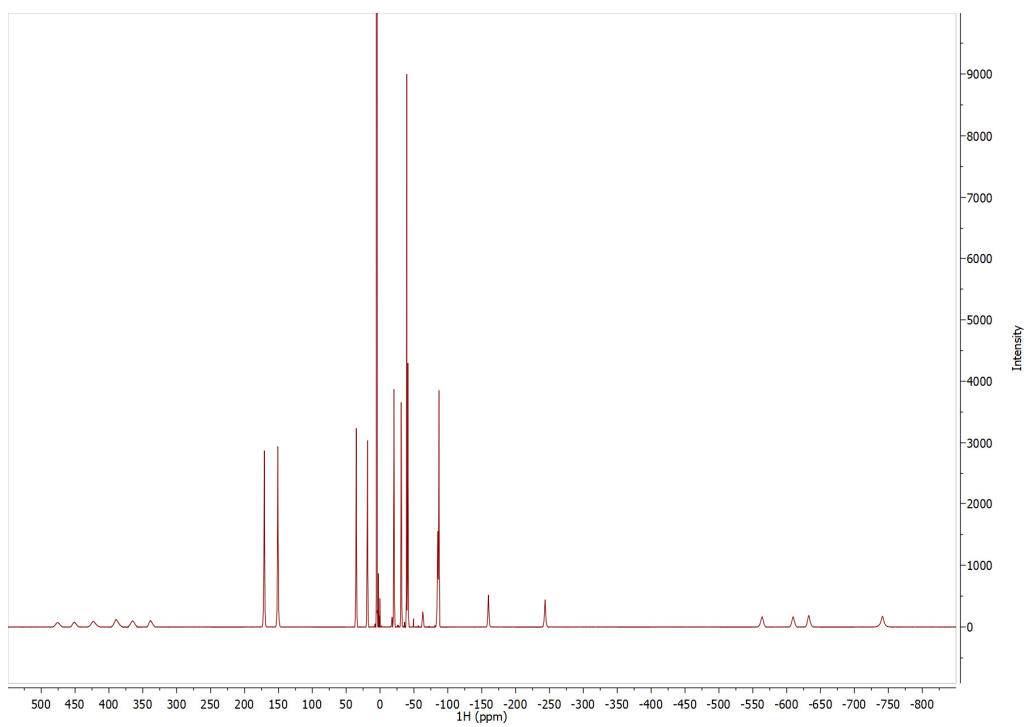


Figure S3: ¹H NMR of Dy-DOTA-M8-ITC (D₂O, 298 K, QCI-F, 600 MHz).

IR spectrum of Tm-DOTA-M8-ITC

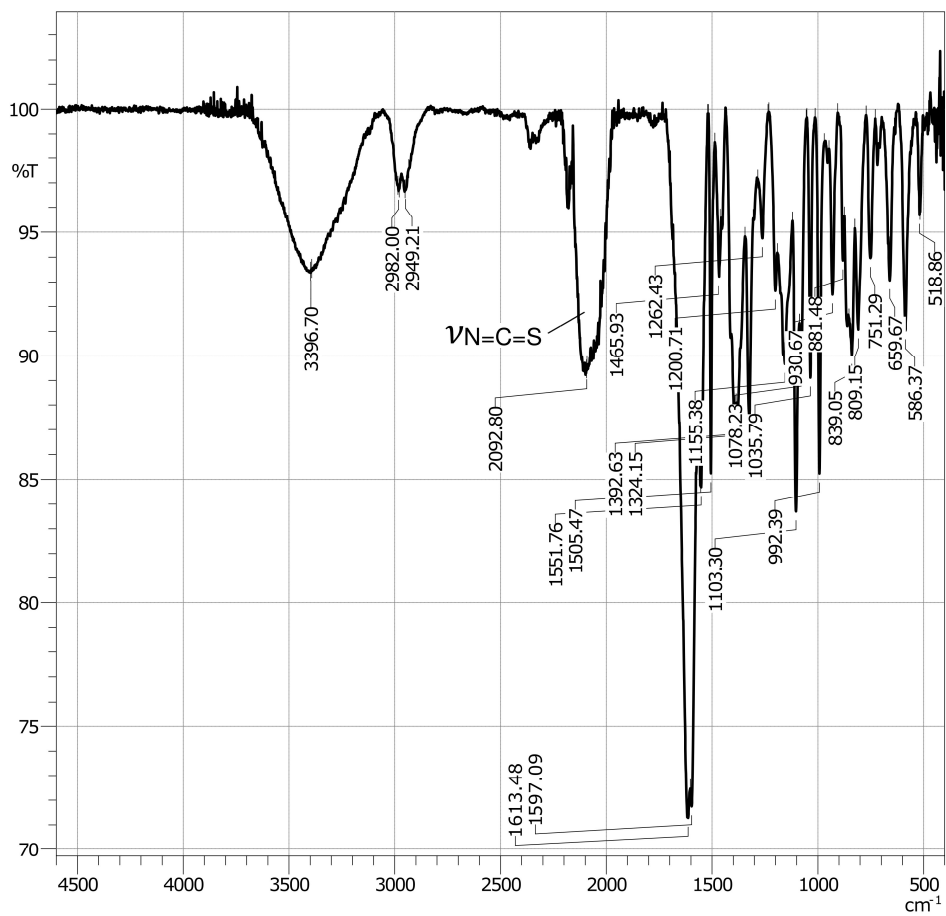


Figure S4: IR spectrum of Tm-DOTA-M8-ITC.

Analytical HPLC-ESI-MS measurement

Lu-DOTA-M8-ITC

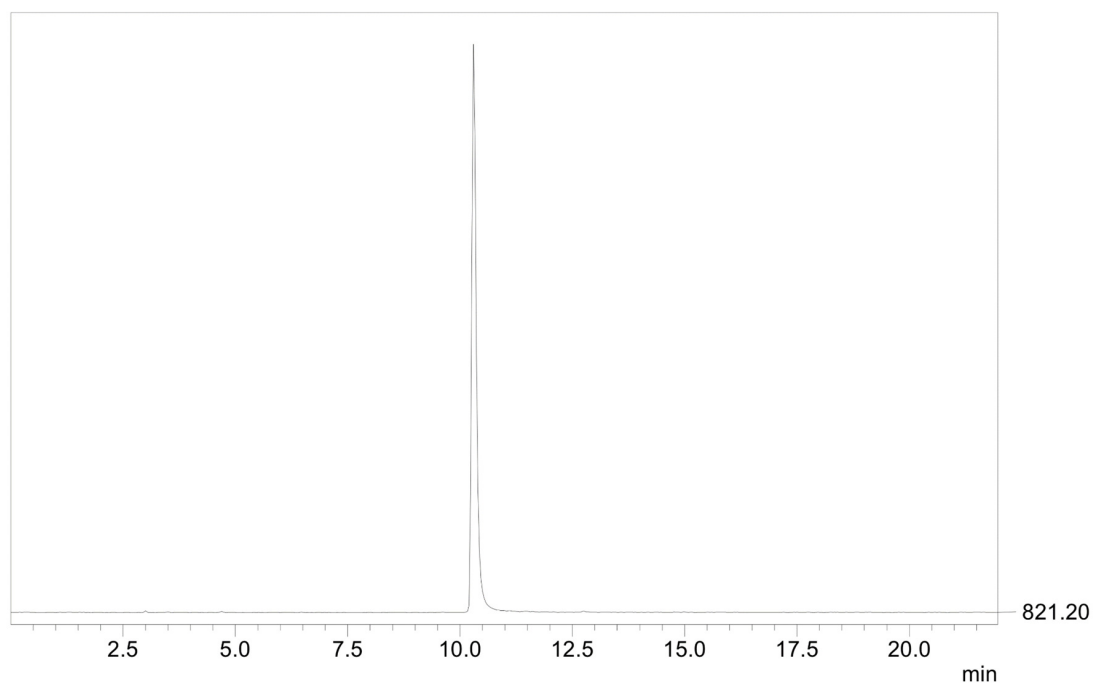


Figure S5: HPLC-ESI-MS trace of Lu-DOTA-M8-ITC.

Tm-DOTA-M8-ITC

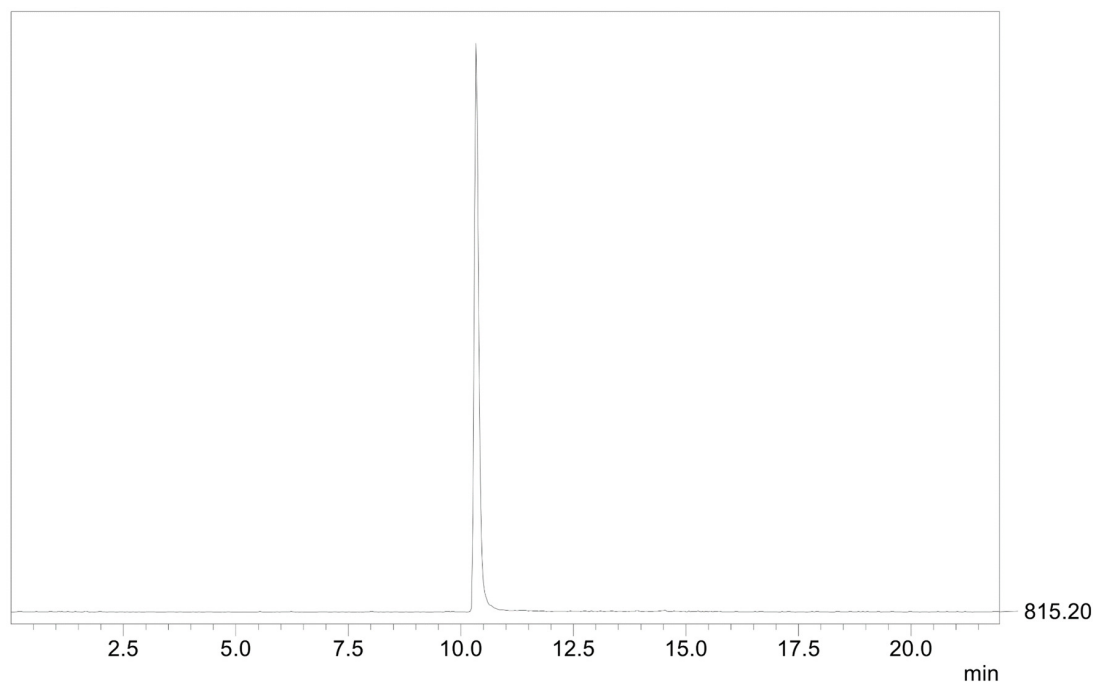


Figure S6: HPLC-ESI-MS trace of Tm-DOTA-M8-ITC.

Dy-DOTA-M8-ITC

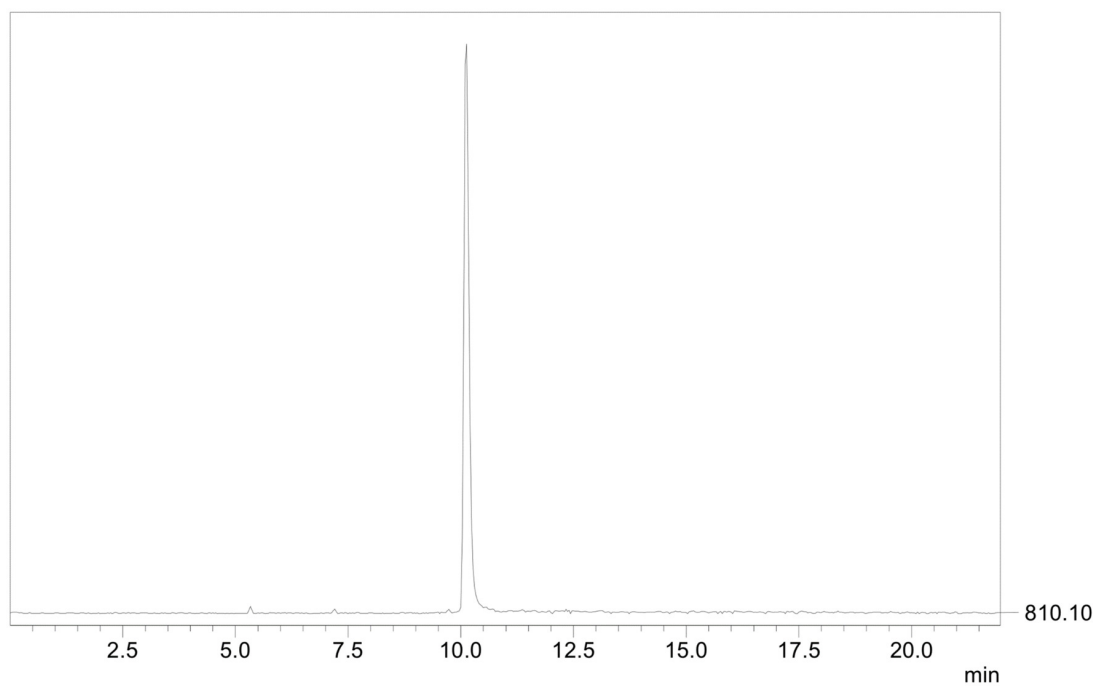


Figure S7: HPLC-ESI-MS trace of Dy-DOTA-M8-ITC.

Overlay of ^1H - ^{15}N SOFAST HMQC spectra of Tm- (red) and Lu-DOTA-M8-ITC (black) labeled SLC^A

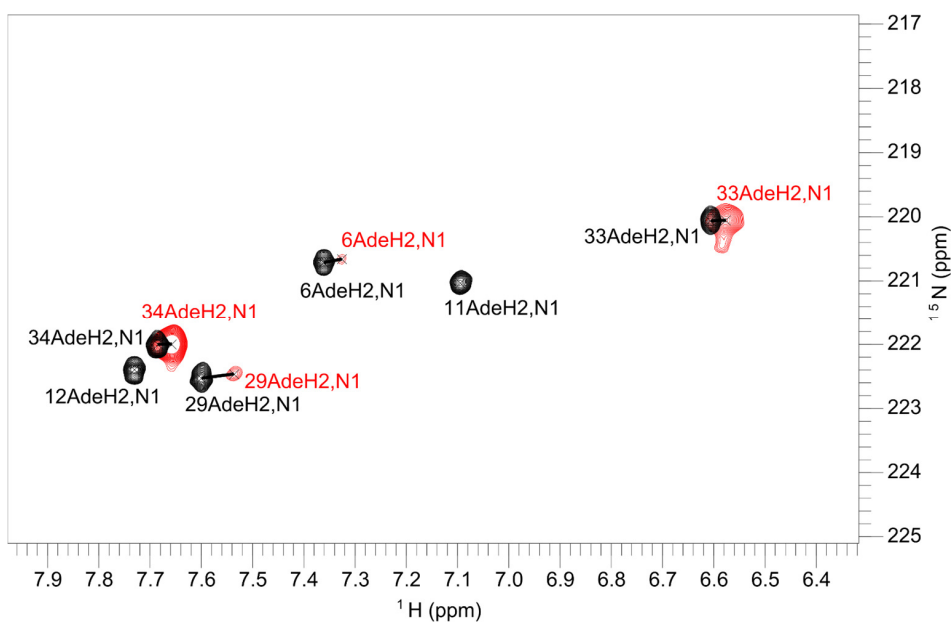


Figure S8: Overlay of ^1H - ^{15}N SOFAST HMQC spectra of Tm- (red) and Lu-DOTA-M8-ITC (black) labeled SLC^A.

Overlay of HNN-COSY spectra of Tm- (red) and Lu-DOTA-M8-ITC (black) labeled SLC^A

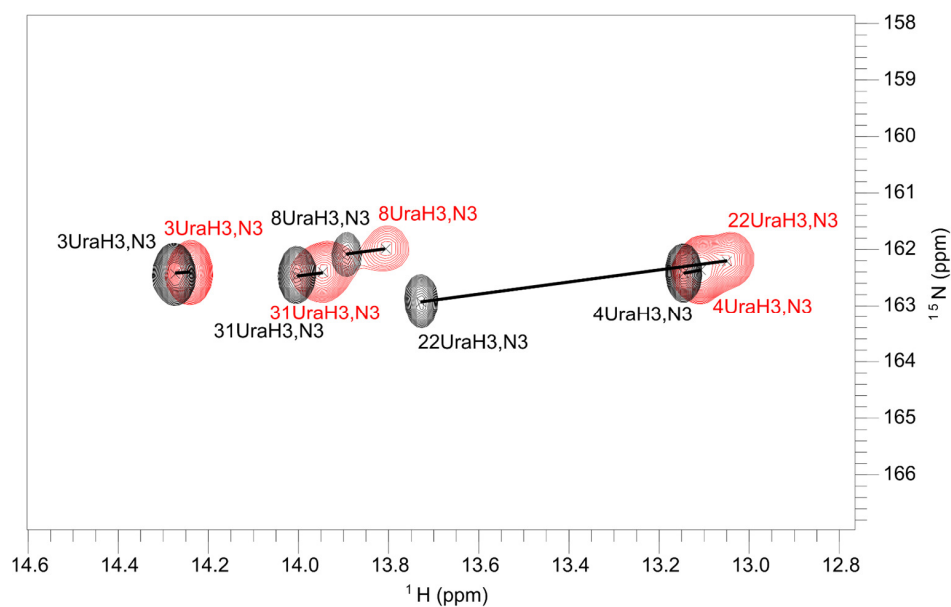


Figure S9: Overlay of HNN-COSY spectra of Tm- (red) and Lu-DOTA-M8-ITC (black) labeled SLC^A.

Overlay of ^1H - ^{13}C TROSY HSQC spectra of Tm- (red) and Lu-DOTA-M8-ITC (black) labeled SLCA^A

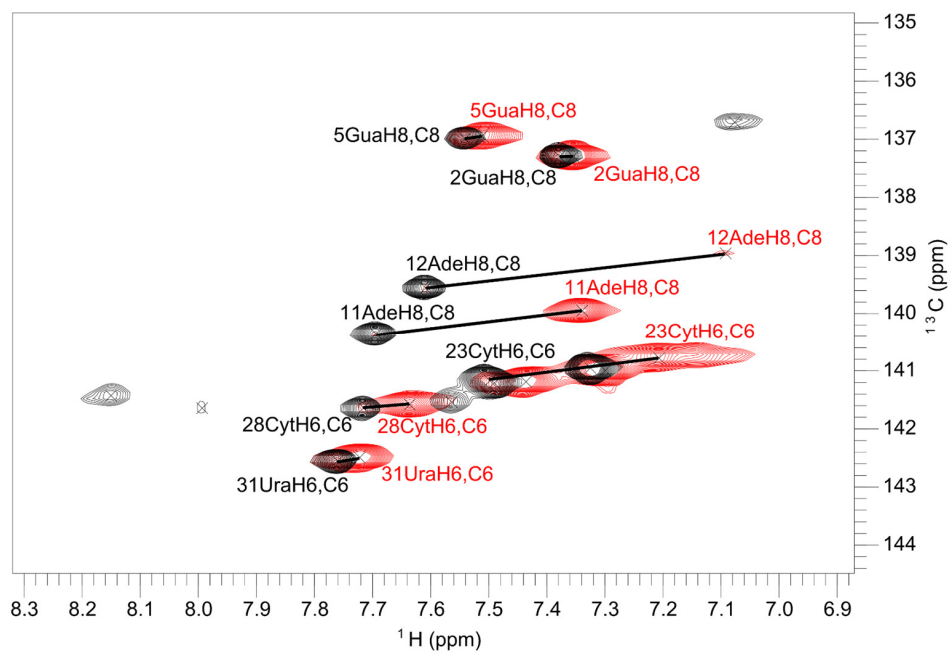


Figure S10: Overlay of ^1H - ^{13}C TROSY HSQC spectra of Tm- (red) and Lu-DOTA-M8-ITC (black) labeled SLCA^A.

Shift list comparison of ^1H - ^{15}N and ^1H - ^{13}C HMQC/HSQC spectra of Tm- and Lu-DOTA-M8-ITC labeled SLC^A

Table S1: Shift list comparison of ^1H - ^{15}N and ^1H - ^{13}C HMQC/HSQC spectra of Tm- and Lu-DOTA-M8-ITC labeled SLC^A.

Base	Atom	PCS
2	C8	-0.003
2	H8	-0.024
3	H3	-0.034
3	N3	-0.012
4	H3	-0.039
4	N3	-0.057
5	C8	-0.043
5	H8	-0.029
6	H2	-0.036
6	N1	-0.053
8	H3	-0.086
8	N3	-0.098
11	C8	-0.404
11	H8	-0.355
12	C8	-0.596
12	H8	-0.517
22	H3	-0.678
22	N3	-0.723
23	C6	-0.388
23	H6	-0.291
28	C6	-0.060
28	H6	-0.080
29	H2	-0.064
29	N1	-0.064
30	C6	-0.021
30	H6	-0.055
31	C6	-0.057
31	H3	-0.058
31	H6	-0.038
31	N3	-0.065
33	H2	-0.030
33	N1	-0.010
34	H2	-0.026
34	N1	0.000
35	C6	0.043
35	H6	-0.026

Correlation plots of experimental and back-calculated PCS

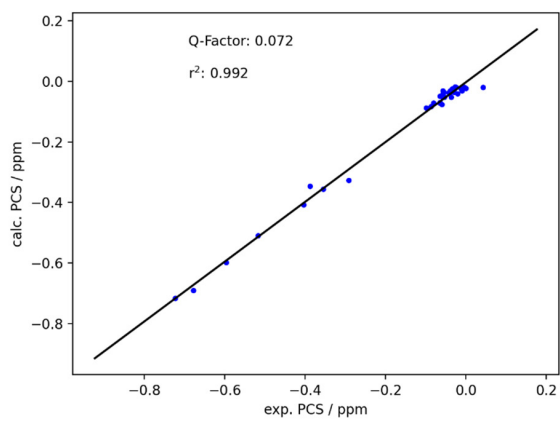


Figure S11: PCS correlation plot of Tm-DOTA-M8-ITC labeled SLC^A.

Overlay of ^1H - ^{13}C TROSY HSQC spectra of apo (black) and holo (labeled with Lu-DOTA-M8-ITC, red) SLC^A

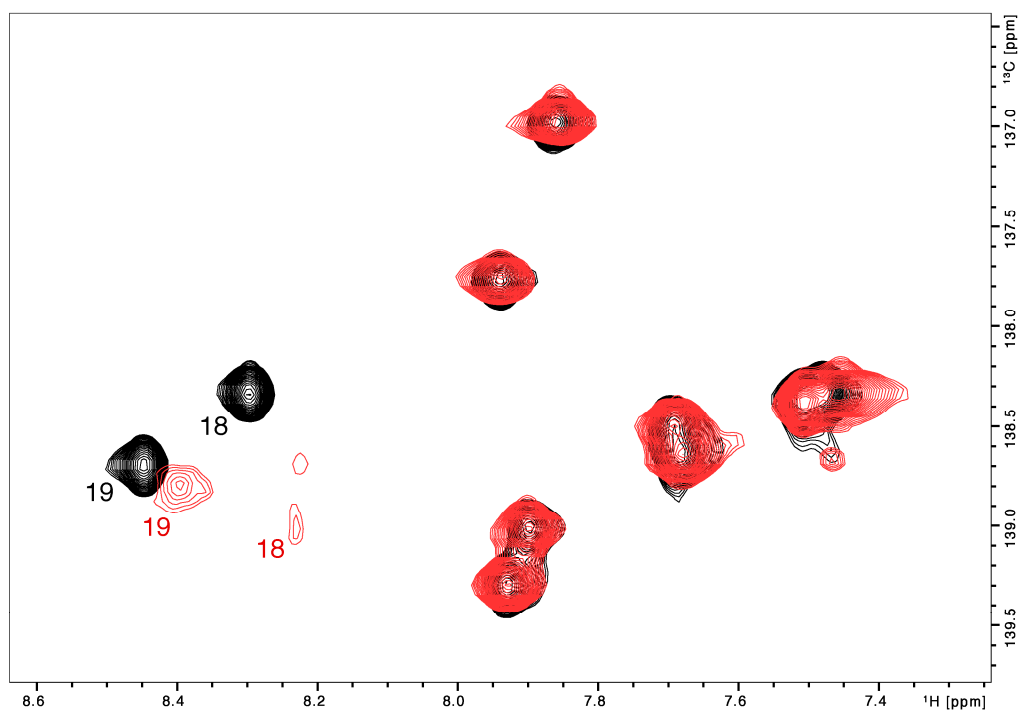


Figure S12: Preliminary experiments performed on an SLC^A RNA with reduced ^{13}C labeling pattern. Overlay of ^1H - ^{13}C TROSY HSQC spectra of apo (black) and holo (labeled with Lu-DOTA-M8-ITC, red) SLC^A. Slight shift differences are observed for the residues close to the tagging site at position 16.

ESI-MS measurements of tagging reactions

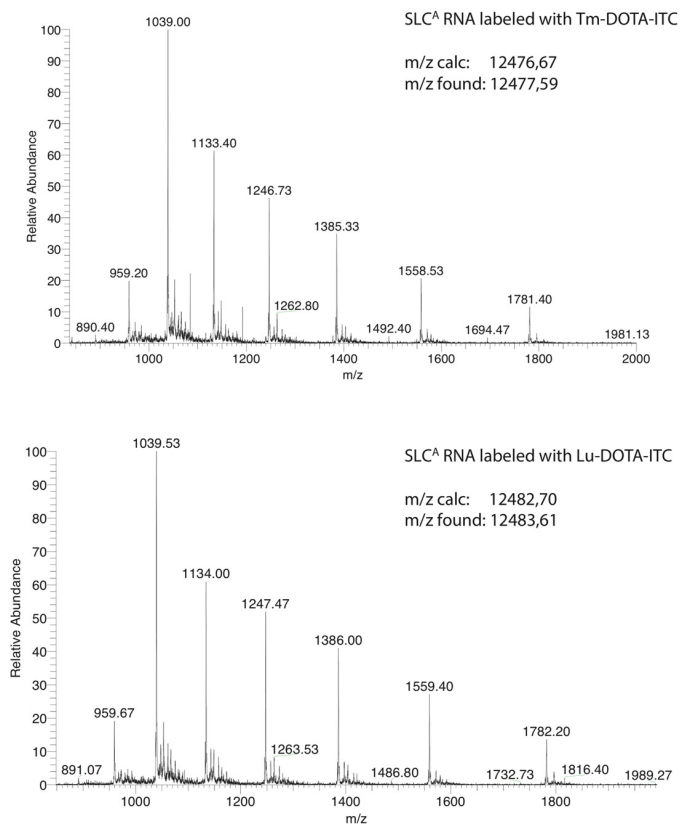


Figure S13: Confirmation of tagging reaction of selectively ¹³C labeled SLC^A. RNA conjugated to Tm- and Lu-DOTA-M8-ITC.

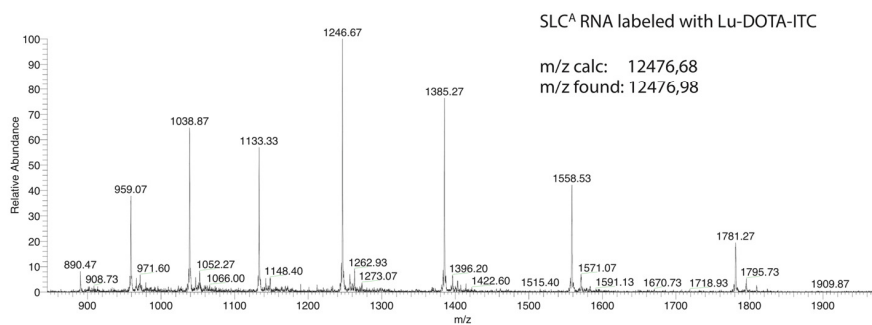
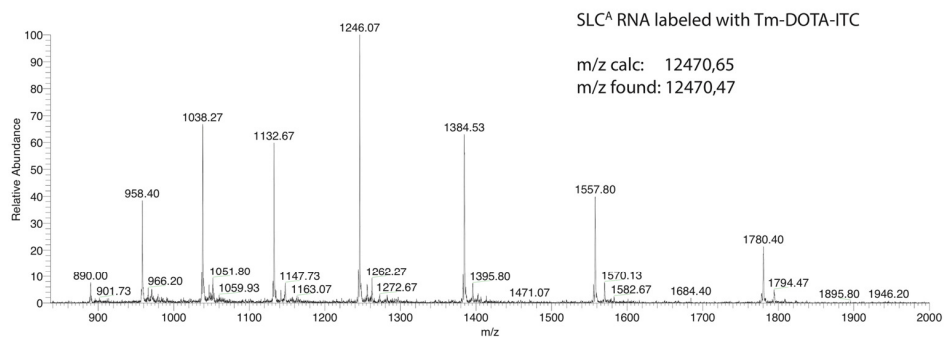


Figure S14: Confirmation of tagging reaction of selectively ¹⁵N labeled SLC^A. RNA conjugated to Tm- and Lu-DOTA-M8-ITC.

Estimation of anisotropy parameters for Tm-DOTA-M8-ITC

Table S2: Estimated anisotropy parameters on the free tag (Tm-DOTA-M8-ITC).

	$\Delta\chi_{ax}$ (10^{-32} m ³)	$\Delta\chi_{rh}$ (10^{-32} m ³)
Tm solvent	95.2	2.6
Tm	104.1	8.4
Average	99.6	5.5

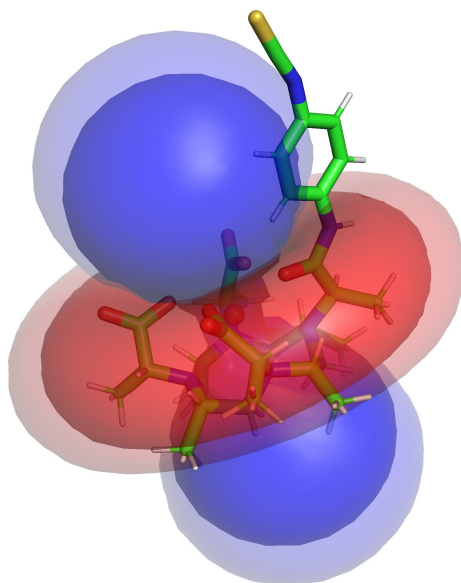


Figure S15: Anisotropy parameters for Tm-DOTA-M8-ITC visualized as isosurfaces with a given PCS (blue: positive PCS, red: negative PCS, inner layer: 80 ppm, outer layer: 50 ppm).

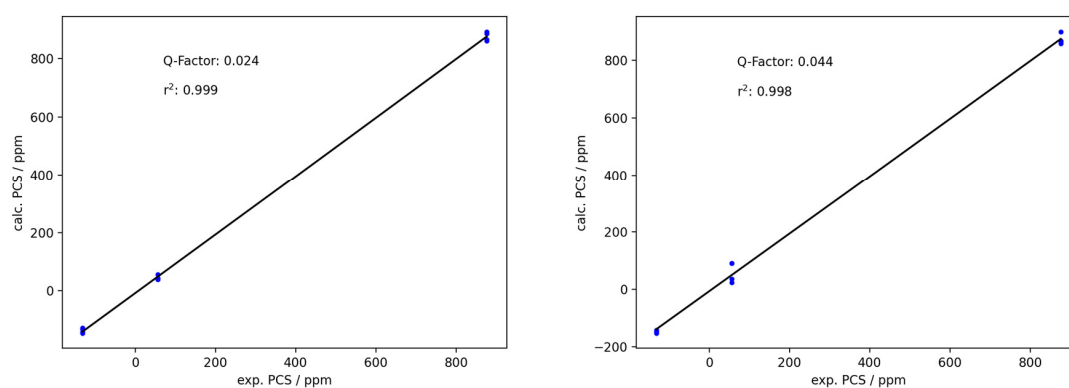


Figure S16: Correlation plots for measured and back-calculated PCS (Tm solvent: left, Tm: right).

Estimation of anisotropy parameters for Dy-DOTA-M8-ITC

Table S3: Estimated anisotropy parameters on the free tag (Dy-DOTA-M8-ITC).

	$\Delta\chi_{ax} (10^{-32} \text{ m}^3)$	$\Delta\chi_{rh} (10^{-32} \text{ m}^3)$
Dy solvent	-69.8	-1.0
Dy	-77.5	-10.4
Average	-73.7	-5.7

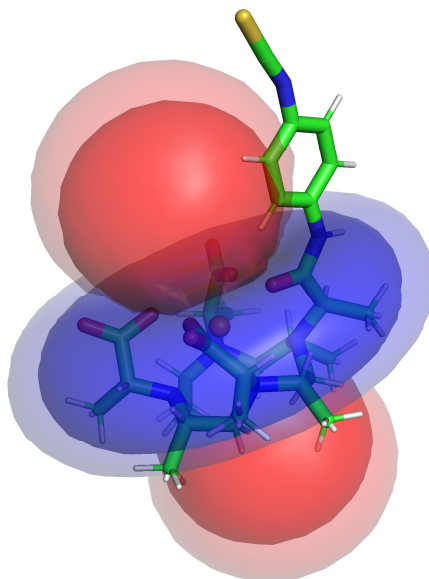


Figure S17: Anisotropy parameters for Dy-DOTA-M8-ITC visualized as isosurfaces with a given PCS (blue: positive PCS, red: negative PCS, inner layer: 80 ppm, outer layer: 50 ppm).

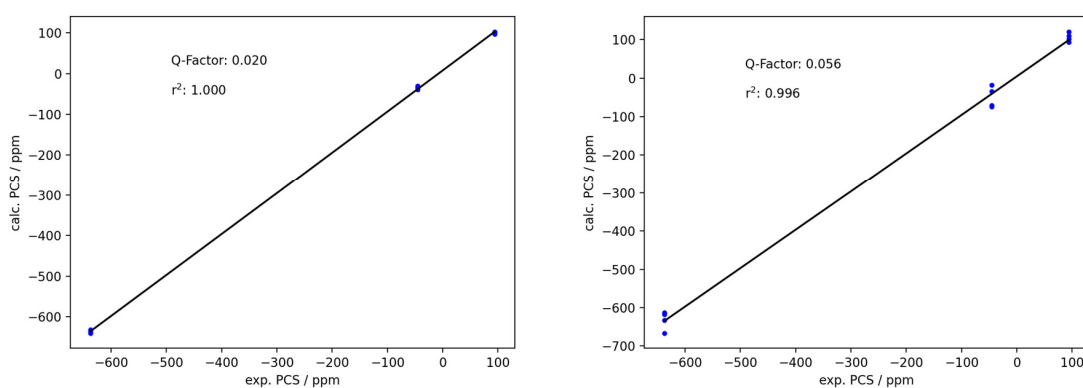


Figure S18: Correlation plots for measured and back-calculated PCS (Dy solvent: left, Dy: right).

DFT calculations

For the calculations, BP86 was used as functional,³⁷⁻³⁸ SARC-TZVP as basis set for the ligand,³⁹ while SARC2-QZVP was used as basis set for the lanthanide metal.⁴⁰ The calculations were performed using the relativistic ZORA approximation,⁴¹ as well as the RI approximation to speed up the calculations.⁴²⁻⁴³ To model the water solvent, CPCM solvent model was implemented into the calculations.⁴⁴

References

1. K. Wüthrich, *Angew. Chem. Int. Ed.*, **2003**, *42*, 3340-3363.
2. B. Alberts, Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P., *Molecular Biology of the Cell*, 4th edition. Garland Science: New York, 2002.
3. D. Joss, D. Häussinger, *Prog. Nucl. Magn. Reson. Spectrosc.*, **2019**, *114-115*, 284-312.
4. C. Nitsche, G. Otting, *Prog. Nucl. Magn. Reson. Spectrosc.*, **2017**, *98-99*, 20-49.
5. G. M. Clore, J. Iwahara, *Chem. Rev.*, **2009**, *109*, 4108-4139.
6. G. M. Clore, C. Tang, J. Iwahara, *Curr. Opin. Struct. Biol.*, **2007**, *17*, 603-616.
7. M. A. S. Hass, M. Ubbink, *Curr. Opin. Struct. Biol.*, **2014**, *24*, 45-53.
8. T. Saio, K. Ogura, H. Kumeta, Y. Kobashigawa, K. Shimizu, M. Yokochi, K. Kodama, H. Yamaguchi, H. Tsujishita, F. Inagaki, *Sci. Rep.*, **2015**, *5*, 16685.
9. C. Göbl, T. Madl, B. Simon, M. Sattler, *Prog. Nucl. Magn. Reson. Spectrosc.*, **2014**, *80*, 26-63.
10. A. Canales, I. Boos, L. Perkams, L. Karst, T. Lubber, T. Karagiannis, G. Domínguez, F. J. Cañada, J. Pérez-Castells, D. Häussinger, C. Unverzagt, J. Jiménez-Barbero, *Angew. Chem. Int. Ed.*, **2017**, *56*, 14987-14991.
11. K. D. Brewer, T. Bacaj, A. Cavalli, C. Camilloni, J. D. Swarbrick, J. Liu, A. Zhou, P. Zhou, N. Barlow, J. Xu, A. B. Seven, E. A. Prinslow, R. Voleti, D. Häussinger, A. M. J. J. Bonvin, D. R. Tomchick, M. Vendruscolo, B. Graham, T. C. Südhof, J. Rizo, *Nat. Struct. Mol. Biol.*, **2015**, *22*, 555-564.
12. K. Zimmermann, D. Joss, T. Müntener, E. S. Nogueira, M. Schäfer, L. Knörr, F. W. Monnard, D. Häussinger, *Chem. Sci.*, **2019**, *10*, 5064-5072.
13. J.-Y. Guan, P. H. J. Keizers, W.-M. Liu, F. Löhr, S. P. Skinner, E. A. Heeneman, H. Schwalbe, M. Ubbink, G. Siegal, *J. Am. Chem. Soc.*, **2013**, *135*, 5859-5868.
14. G. Pintacuda, M. John, X.-C. Su, G. Otting, *Acc. Chem. Res.*, **2007**, *40*, 206-212.
15. D. Joss, R. Vogel, D. Häussinger, *Ref. Mod. in Chem, Mol. Sci. and Chem. Eng.*, **2020**, doi:10.1016/B978-0-12-409547-2.14848-6.
16. D. Joss, M.-S. Bertrams, D. Häussinger, *Chem. Eur. J.*, **2019**, *25*, 11910-11917.
17. D. Joss, D. Häussinger, *Chem. Commun.*, **2019**, *55*, 10543-10546.
18. P. H. Keizers, A. Saragliadis, Y. Hiruma, M. Overhand, M. Ubbink, *J. Am. Chem. Soc.*, **2008**, *130*, 14802-14812.
19. B. Graham, C. T. Loh, J. D. Swarbrick, P. Ung, J. Shin, H. Yagi, X. Jia, S. Chhabra, N. Barlow, G. Pintacuda, T. Huber, G. Otting, *Bioconjugate Chem.*, **2011**, *22*, 2118-2125.
20. F. Peters, M. Maestre-Martinez, A. Leonov, L. Kovacic, S. Becker, R. Boelens, C. Griesinger, *J. Biomol. NMR*, **2011**, *51*, 329-337.
21. J. Koehler, J. Meiler, *Prog. Nucl. Magn. Reson. Spectrosc.*, **2011**, *59*, 360-389.
22. K. Jaudzems, X. Jia, H. Yagi, D. Zhulenkova, B. Graham, G. Otting, E. Liepinsh, *J. Mol. Biol.*, **2012**, *424*, 42-53.
23. M. Strickland, J. Catazaro, R. Rajasekaran, M.-P. Strub, C. O'Hern, G. A. Bermejo, M. F. Summers, J. Marchant, N. Tjandra, *J. Am. Chem. Soc.*, **2019**, *141*, 1430-1434.
24. Z. Wu, M. D. Lee, T. J. Carruthers, M. Szabo, M. L. Dennis, J. D. Swarbrick, B. Graham, G. Otting, *Bioconjug. Chem.*, **2017**, *28*, 1741-1748.
25. S. Täubert, Y.-H. Zhang, M. M. Martinez, F. Siepel, E. Wöltjen, A. Leonov, C. Griesinger, *ChemBioChem*, **2020**, *n/a*.
26. J. Kremser, E. Strebiter, R. Plangger, M. A. Juen, F. Nußbaumer, H. Glasner, K. Breuker, C. Kreutz, *Chem. Commun.*, **2017**, *53*, 12938-12941.
27. S. Neuner, C. Kreutz, R. Micura, *Monatsh. Chem.*, **2017**, *148*, 149-155.
28. S. Neuner, T. Santner, C. Kreutz, R. Micura, *Chem. Eur. J.*, **2015**, *21*, 11634-11643.
29. C. H. Wunderlich, R. Spitzer, T. Santner, K. Fauster, M. Tollinger, C. Kreutz, *J. Am. Chem. Soc.*, **2012**, *134*, 7558-7569.
30. D. Häussinger, J.-R. Huang, S. Grzesiek, *J. Am. Chem. Soc.*, **2009**, *131*, 14761-14767.
31. D. Joss, R. M. Walliser, K. Zimmermann, D. Häussinger, *J. Biomol. NMR*, **2018**, *29*-38.
32. J. Marchant, A. Bax, M. F. Summers, *J. Am. Chem. Soc.*, **2018**, *140*, 6978-6983.
33. T. Müntener, J. Kottelat, A. Huber, D. Häussinger, *Bioconjug. Chem.*, **2018**, *29*, 3344-3351.
34. D. Joss, F. Winter, D. Häussinger, *Chem. Commun.*, **2020**, *56*, 12861-12864.
35. W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides, E. D. Laue, *Proteins*, **2005**, *59*, 687-696.
36. C. Schmitz, M. J. Stanton-Cook, X.-C. Su, G. Otting, T. Huber, *J. Biomol. NMR*, **2008**, *41*, 179.
37. A. D. Becke, *Phys. Rev. A*, **1988**, *38*, 3098-3100.
38. J. P. Perdew, *Phys. Rev. B*, **1986**, *33*, 8822-8824.
39. D. A. Pantazis, F. Neese, *J. Chem. Theory Comput.*, **2009**, *5*, 2229-2238.

40. D. Aravena, F. Neese, D. A. Pantazis, *J. Chem. Theory Comput.*, **2016**, *12*, 1148-1156.
41. E. van Lenthe, E. J. Baerends, J. G. Snijders, *J. Chem. Phys.*, **1993**, *99*, 4597-4610.
42. R. A. Kendall, H. A. Früchtl, *Theor. Chem. Acc.*, **1997**, *97*, 158-163.
43. O. Vahtras, J. Almlöf, M. W. Feyereisen, *Chem. Phys. Lett.*, **1993**, *213*, 514-518.
44. Y. Takano, K. N. Houk, *J. Chem. Theory Comput.*, **2005**, *1*, 70-77.

6 Part IV

6.1 **Structure of formylglycine-generating enzyme in complex with copper and a substrate reveals an acidic pocket for binding and activation of molecular oxygen**

Reproduced from D. A. Miarzlou, F. Leisinger, D. Joss, D. Häussinger, F. P. Seebeck, *Chemical Science*, **2019**, 10, 7049-7058 with permission from the Royal Society of Chemistry.

Cite this: *Chem. Sci.*, 2019, 10, 7049

All publication charges for this article have been paid for by the Royal Society of Chemistry

Structure of formylglycine-generating enzyme in complex with copper and a substrate reveals an acidic pocket for binding and activation of molecular oxygen†

Dzmitry A. Miarzlou,^{ID} Florian Leisinger,^{ID} Daniel Joss,^{ID} Daniel Häussinger^{ID} and Florian P. Seebeck^{ID}*

The formylglycine generating enzyme (FGE) catalyzes oxidative conversion of specific peptidyl-cysteine residues to formylglycine. FGE mediates O₂-activation and hydrogen-atom abstraction in an active site that contains Cu(I) coordinated to two cysteine residues. Similar coordination geometries are common among copper-sensing transcription factors and copper-chaperone but are unprecedented among copper-dependent oxidases. To examine the mechanism of this unusual catalyst we determined the 1.04 Å structure of FGE from *Thermomonospora curvata* in complex with copper and a cysteine-containing peptide substrate. This structure unveils a network of four crystallographic waters and two active site residues that form a highly acidic O₂-binding pocket juxtaposed to the trigonal planar tris-cysteine coordinated Cu(I) center. Comparison with structures of FGE in complex with Ag(I) and Cd(II) combined with evidence from NMR spectroscopy and kinetic observations highlight several structural changes that are induced by substrate binding and prime the enzyme for O₂-binding and subsequent activation.

Received 8th April 2019
Accepted 11th June 2019DOI: 10.1039/c9sc01723b
rsc.li/chemical-science

Introduction

Copper is a versatile catalyst for oxygen-dependent reactions. Combined with appropriate ligands, copper can cycle between the oxidation states I, II and III to activate molecular oxygen (O₂) and to form reactive oxygen species that can initiate very difficult reactions.^{1–3} Mechanistic appreciation of this reactivity is an important prerequisite to understand the fundamental role of copper in biology and to exploit its potential in chemical synthesis.⁴ There are two groups of copper-proteins that are most relevant for the subject of the present report: mononuclear copper enzymes that utilize O₂ as an oxidant, and copper-binding proteins that suppress unspecific and cytotoxic activities of copper in the cell.⁵

The group of known mononuclear copper enzymes includes lytic polysaccharide monoxygenases (LPMO),^{6,7} particulate methane monoxygenase (pMMO),⁸ peptidylglycine- α -hydroxylating monoxygenase (PHM),^{9,10} copper amine oxidase (CAO),¹¹ and dopamine- β -hydroxylase (DBH).¹² These enzymes bind copper in histidine-dominated tetrahedral or square planar coordination spheres. The structure of these enzymes have inspired the development of abiotic copper-ligands in

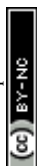
order to examine specific aspects of copper-catalysis, but also with an eye on application in organic synthesis.^{2,13} Not surprisingly, most of these compounds also contain nitrogen-rich ligand sets.

The mononuclear copper proteins of the second group do not necessarily catalyze reactions but rather contribute to cellular copper-transport and sensing. Representative examples are the Cu-responsive transcriptional activator CueR,¹⁴ and the copper chaperone Atox1.¹⁵ These proteins bind Cu(I) in a linear bis-cysteine coordination sphere with extremely high affinity.¹⁶ This coordination-type provides an ideal solution to form redox inert protein:copper complexes with maximal thermodynamic stability, while enabling reversible transfer of Cu(I) from one protein to another.^{15,17}

The formylglycine generating enzyme (FGE) is unique, since it cannot be attributed to either the first or the second group of copper-binding proteins. FGE binds a single Cu(I) by linear bis-cysteine coordination and catalyzes copper-dependent O₂ activation.^{18,19} FGE catalyzes the oxidative conversion of specific peptidyl-cysteines to formylglycine *via* abstraction of the pro-(*R*)- β -hydrogen atom (fGly, Fig. 1). This posttranslational modification is important for sulfatases which use the hydrated form of fGly as catalytic nucleophile.²⁰ FGE has attracted scientific interest for several reasons. Initially, this enzyme was discovered in the pursuit of finding the molecular cause for the storage disease *multiple sulfatase deficiency*.^{20,21} The subsequent realization that FGE can be used as a tool to introduce

Department of Chemistry, University of Basel, Mattenstrasse 24a, Basel 4002, Switzerland. E-mail: florian.seebeck@unibas.ch; Tel: +41 612071143

† Electronic supplementary information (ESI) available: Detailed descriptions of all experiments, supporting Fig. (S1–S15) and Tables (S1–S4) are shown in the ESI. See DOI: 10.1039/c9sc01723b



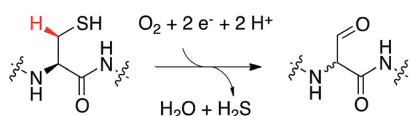


Fig. 1 FGE-catalyzed oxidation of peptidyl-cysteine to formylglycine (fGly) is initiated by abstraction of the pro-(*R*)- β -hydrogen atom from the substrate (red).

electrophilic aldehyde functions into recombinant proteins highlighted the considerable biotechnological potential of this enzyme.^{22–24} Finally, the recent discovery that FGE is a copper-dependent enzyme opened the door for mechanistic studies on this unique enzyme.^{18,19,25,26} The crystal structure of FGE from *Thermomonospora curvata* (*tcFGE*) in complex with Ag(I) (*tcFGE_Ag*, PDB: 5NXL), and most recently, the crystal structure of FGE from *Streptomyces coelicolor* in complex with Cu(I) (*scFGE_Cu*, PDB: 6MUJ), combined with biochemical characterization showed that two active site cysteines are the only metal ligands and that the Cu(I) bound state is the catalytic resting state.^{25–27} In this way, FGE represents an entirely new type of copper-enzyme. Structural and mechanistic studies on this enzyme may inspire new directions in the design of abiotic copper-catalysts.² Conversely, the similarities of the primary metal coordination sphere in FGE and non-catalytic Cu-binding proteins raise the questions as to what secondary features may enable O₂-activation by FGE. Identification of these features may help to identify redox enzymes, masquerading as copper-trafficking proteins, or could provide guidelines to engineer novel redox-activity into existing copper-proteins.

In this report we describe the crystal structure of *tcFGE* in complex with Cu(I) and a 17-residue substrate analog. Based on this structure together with NMR spectroscopy and kinetic characterization, we identify an acidic O₂-binding pocket juxtaposed to the copper center as key determinant for efficient O₂-activation.

Results and discussion

Crystallization of *tcFGE* in complex with Cu(I) and a substrate-analog

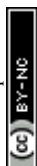
Recombinant *tcFGE* was produced in *Escherichia coli* (see ESI†).¹⁹ The purified enzyme was crystallized in the presence of 1.2 equivalents of Cu(I) and 2.8 equivalents of a synthetic 17-residue substrate-analog. To stabilize the enzyme:substrate complex, crystallization and room-temperature handling of the crystals were executed under anoxic conditions. The sequence of the substrate-analog (FGE-27: Abz-ATTPLCGPSRASILSGR, Abz = *o*-aminobenzoic acid) was derived from the active site segment of a putative arylsulfatase (WP_012850446) from *T. curvata*. Using isothermal titration calorimetry (ITC), we established that this peptide binds to an inactive variant of *tcFGE* (*tcFGE*_{W228F}, see below) with a $K_d = 40 \pm 10$ nM (Fig. S2†), indicating that the wild type enzyme would also form a stable complex with this peptide. Single crystals were isolated from solutions containing purified *tcFGE* (21.5 mg mL⁻¹), CuCl, peptide, PEG 8000 (10%), Tris-HCl (0.1 M, pH 7.0) and MgCl₂

(0.2 M). Diffraction data could be processed up to a resolution of 1.04 Å (Fig. 2A). At this resolution conservative estimates of standard deviations (r.m.s.) are ± 0.05 Å for bond lengths and $\pm 2^\circ$ for bond angles.²⁸ The unbiased electron density revealed a continuous polypeptide chain from –1 to 302 for *tcFGE* and from Abz (–6) to Gly (11) for the substrate FGE-27. Two Ca(II) ions were identified in two conserved calcium binding sites (Fig. S3†).^{26,29,30} The high resolution of this data set allowed placement of 494 crystallographic waters, in particular four waters in the active site (H₂O_1–4) with high confidence (Fig. S4†). The overall structure of *tcFGE* and its similarity to human FGE (*hsFGE*) and *scFGE* has been discussed previously.^{26,29,30} Therefore, the main focus here will be on the structure of the substrate binding pocket and the metal binding site. For this discussion we compare *tcFGE* in complex with Cu(I) and substrate (*tcFGE_Cu_S*) with the structure of *tcFGE* in complex Ag(I) (*tcFGE_Ag*) as a model for *tcFGE* in complex with Cu(I) (*tcFGE_Cu*). The structure of *scFGE* in complex with Cu (*scFGE_Cu*) has also become available very recently.²⁷ The metal binding sites in *scFGE_Cu* and *tcFGE_Ag* were found to be essentially isostructural.

Position of the substrate

The unbiased electron density for the peptide reveals ordered positions for all residues except for the C-terminal Arg10 (Fig. 2B). The residues in the N-terminal part (Ala-5–Ser4), except for Cys1, adopt extended backbone conformations ($\phi = -50^\circ$ to -136° ; $\psi = 120$ – 170°). The conformation of Cys1 ($\phi = -108^\circ$; $\psi = -20^\circ$) induces a sharp turn allowing this residue to access the copper localized at the deepest point of the substrate binding pocket (Fig. 2). Pro-2 and Pro-3 make extensive hydrophobic interactions with Phe38/Tyr273 and Phe86/Trp109 respectively. In contrast, Ala-5, Thr-4, Thr-3 and Leu-1 make little or no contact with the protein surface. The C-terminal residues Arg3–Ile6 adopt dihedral angles consistent with a helical conformation ($\phi = -90$ to 62 ; $\psi = -30$ to -5). Indeed, the network of intramolecular backbone-backbone hydrogen bonds is reminiscent of a short 3_{10} -helix (Fig. 2C). The side chain of Ser4 hydrogen bonds to the backbone amide of Ser7 (3.0 Å), serving as a N-terminal helix cap.³¹ In return, the side chain of Ser7 hydrogen bonds with the backbone carbonyl of Ser4 (2.7 Å). The guanidinium side chain of Arg5 makes a π -cation interaction with the indole ring of Trp84, in addition to hydrogen-bonds with Ser290 (2.8 Å), and the backbone carbonyl of Trp84 (2.9 Å). The side chain of Ile8 binds to a hydrophobic dell lined by the side chains of Trp84, Phe86, Met99 and Ala101. The side chains of Ala6, Leu9 and the following residues (Ser10, Gly11 and Arg12) make no interaction with the protein surface. The helical structure of the substrate may be further stabilized by weak interactions with neighboring *tcFGE* chains (Fig. 2D and E).

hsFGE has been crystallized with a seven-residue substrate (LCTPSRA, *hsFGE_S*, PDB: 2AIK).³² This structure contains no metal in the active site and Cys366 (Cys274 in *tcFGE*) was mutated to Ser. Instead, the substrate cysteine is linked *via* disulfide bond to Cys341 (Cys269 in *tcFGE*). Despite these damages, the equivalent substrate residues in *tcFGE_Cu_S* and



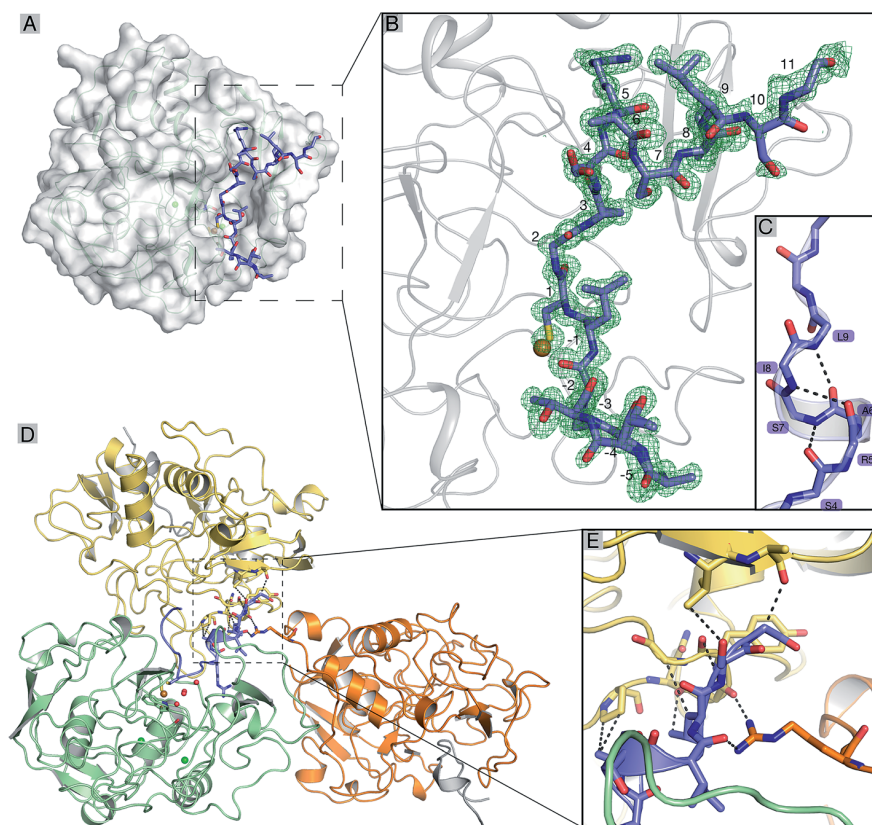


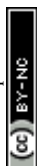
Fig. 2 (A) Structure of *tcFGE* in complex with Cu(I) and substrate (*tcFGE_Cu_S*). The 17-residue peptide substrate (violet). (B) The $m|F_o| - D|F_c|$ omit map for the substrate model is contoured at σ -level = 3.0. (C) The C-terminus of the substrate folds into a short 3_{10} -helix and interacts with neighboring *tcFGE* chains in the crystal (D and E). The dashed lines in (E) indicate nearest contacts, not necessarily attractive interactions.

hsFGE_S adopt almost identical positions in the substrate binding pocket (Fig. S5†). Apparently, residues 0–5, the conserved residues Pro3 and Arg5 in particular (numbering according to *tcFGE_Cu_S*) make sufficient interactions with the substrate binding pocket to steer the targeted cysteine residue into the active site – even in the absence of Cu(I).

Indeed, the biotechnological potential of FGE depends on its ability to target cysteine residues embedded in the minimal LCXPXR motif. The limited length of this motif can be inserted into N- or C-termini or surface exposed loops of unrelated proteins without significant disturbance of the structure and stability of the parent protein.^{22,23,33,34} On the other hand, comparison of kinetic parameters from *in vitro* studies with 9-, 14- and 23-residue substrates showed that longer substrates are recognized with significantly higher affinity.^{18,19,35} The structure of *tcFGE_Cu_S* provides a convincing explanation for this observation. The C-terminal residues Arg5–Ile8 adopt a defined secondary structure that positions Ile8 for additional hydrophobic contact with FGE. Likewise, Pro-2 also makes a specific contact with the enzyme that has not been observed in the structure of *hsFGE_S*. The more comprehensive picture of the enzyme:substrate interaction seen in the structure of *tcFGE_Cu_S* could provide a valuable starting point for the design of FGE variants that accept alternative motifs with applications in orthogonal bifunctionalization of recombinant proteins.³⁶

Primary and secondary coordination sphere

The structure of *FGE_Cu_S* demonstrates that the peptidyl cysteine (Cys1) from the substrate is a direct ligand of the catalytic Cu(I). This finding corroborates XAS observations suggesting trigonal coordination of Cu(I) by *scFGE* and its substrate.²⁷ The three thiol side chains of Cys269 (Cu–S_{Cys269}: 2.2 Å), Cys274 (Cu–S_{Cys274}: 2.3 Å) and the Cys1 (Cu–S_{Cys1}: 2.2 Å) coordinate Cu(I) in an irregular trigonal planar complex with significant deviations from the expected 120° S–Cu–S bond angles (S_{Cys269}–Cu–S_{Cys274}: 108°; S_{Cys274}–Cu–S_{Cys1}: 104°; S_{Cys269}–Cu–S_{Cys1}: 147°). The angle between the two FGE-derived cysteines and Cu(I) is significantly smaller than the value (134°) derived from DFT-based modelling based on the structure of *scFGE_Cu*.²⁷ Instead, the much larger S_{Cys269}–Cu–S_{Cys1} angle opens the triangle enabling two crystallographic waters (H₂O₁, 3.3 Å; H₂O₂, 3.5 Å) to form near van der Waals contact with Cu(I). Notably, these two waters are also immobilized by planar trigonal solvation spheres. H₂O₂ is in contact with H₂O₁, H₂O₃ and H₂O₄ (all 2.8 Å, O₁–O₂–O₄: 111°; O₁–O₂–O₃: 124°; O₃–O₂–O₄: 119°). H₂O₁ also binds to the indole side chain of Trp228 (2.9 Å) and Ser266 (2.8 Å, O_{Ser266}–O₁–N_{Trp228}: 89°; N_{Trp228}–O₁–O₂: 129°; O₂–O₁–O_{Ser266}: 140°). In addition, H₂O₁ and H₂O₂ make weak contacts with the thiolates of Cys269 (3.1 Å) and Cys-1 (3.3 Å). This solvation geometry deviates significantly from the more stable tetrahedral solvation spheres



observed in bulk water,³⁷ suggesting that the two crystallographic waters are held in unstable positions. Release of H₂O₁ and H₂O₂ from this pocket may provide important enthalpic driving force to enhance O₂-binding (see below).^{38–41}

Inspection of the next neighbors to Ser266, H₂O₃ and H₂O₄ provides information about the position of the protons in this hydrogen-bonding network. In addition to contacting H₂O₁, the side chain of Ser266 also hydrogen bonds with the backbone amide of Leu268 (3.0 Å). This interaction suggests that Ser266 is a hydrogen-bond donor with respect to H₂O₁. H₂O₃ is in hydrogen-bonding contact with Trp108 (2.9 Å), Trp109 (2.9 Å) and the backbone carbonyl of Leu-1 of the substrate (2.7 Å). This tetrahedral system suggests that H₂O₃ is a hydrogen bond donor with respect to H₂O₂. H₂O₄ is in contact with the backbone carbonyl of Leu268 (2.8 Å) and is otherwise engulfed by hydrophobic moieties of Phe194, Tyr273 and Cys269. This environment too suggests that H₂O₄ is a hydrogen bond donor with respect to H₂O₂. Hence, the pocket that hosts H₂O₁ and H₂O₂ in the *tcFGE*_Cu_S structure appears as a highly acidic microenvironment.

Hydrogen bonding to metal-coordinated thiolates are known to affect the nucleophilicity and redox potential of metal-thiolate complexes.^{42–45} Such interactions are indeed observable in *tcFGE*_Cu_S. Cys1 makes two weak hydrogen bonds to H₂O₂ and Arg276, 3.3 Å, Cys269 hydrogen bonds to H₂O₁ and the backbone amide of Ser266 (3.3 Å), whereas Cys274 only points to the backbone amide of His270 (3.6 Å). This network may play an important role in stabilizing the high electron density of the tris-thiolate Cu(i) complex. Incidentally, Cys274 appears as the least stabilized thiolate in the complex suggesting that this residue may be most vulnerable to oxidation during catalysis or as a result of abortive side reactions (see below).

Comparison of Ag, Cu and Cd-bound FGE reveals a sequence of coordination changes

The structure of *tcFGE*_Cu_S (resolution: 1.04 Å) complements the structures of the same enzyme in complex with Ag(i) (resolution: 1.66 Å) or with Cd(ii) (*tcFGE*_Cd, PDB: 5NYY, resolution: 1.28 Å). Comparison of the three structures provide a high-resolution picture of the structural changes that take place as the coordination sphere expands from linear to trigonal to tetrahedral in order to allow inner-sphere coordination of substrate and O₂.

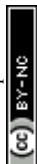
Superposition of the *tcFGE*_Ag and *tcFGE*_Cu_S structures highlights three significant changes that are induced by substrate binding. First, the metal center moves by 1.1 Å as it changes from linear to trigonal coordination (Fig. 3). This movement reduces the distance from the metal to H₂O₁ and H₂O₂ by about 0.3 Å and 0.8 Å respectively. As a second change, H₂O₁ and H₂O₂ are displaced by 0.6 Å from predominantly tetrahedral to purely trigonal planar hydrogen bonding environments. In addition, the interaction partners of H₂O₂ change. In *tcFGE*_Ag H₂O₂ hydrogen bonds with four partners: H₂O₁ (2.6 Å) H₂O₅₀ (equivalent to H₂O₃, 2.8 Å), H₂O₁₀₅ (2.7 Å) and Tyr273 (2.8 Å, Fig. S7†). Substrate binding displaces H₂O₁₀₅, leaving H₂O₂ with only three nearest

neighbors in *tcFGE*_Cu_S. In addition, the side chain of Tyr273 is rotated out of the active site to establish a hydrogen-bond to the backbone amide of Leu-1 of the substrate. Instead, H₂O₄ is recruited as third neighbor of H₂O₂ in *tcFGE*_Cu_S. This exchange is likely accompanied with a change in hydrogen bond polarity. In *tcFGE*_Ag Tyr273 hydrogen bonds with the backbone carbonyl of Leu268 (2.7 Å) (Fig. S7†), indicating that Tyr273 is a hydrogen bond acceptor with respect to H₂O₂. In contrast, H₂O₄ in *tcFGE*_Cu_S likely acts as a hydrogen bond donor (see above). These changes suggest that substrate binding induces destabilization of H₂O₁ and H₂O₂. A third substrate-induced change occurs at Cys274. The dihedral angle along C α -C β of this residue changed by 58° (N α -C α -C β -S) and the amide function between Cys274 and Asn275 rotated by 135° presumably as a result of the change from linear to trigonal Cu(i)-coordination (Fig. S6†). A similar rotation was observed in the *tcFGE*_Cd structure (135°). Hence, this reorganization also seems to be driven by substrate binding.

The structure of *tcFGE*_Cd provides a model for the enzyme with a tetrahedral coordination site which may be required simultaneous coordination of the substrate and a reactive oxygen species.²⁷ Superposition of the structures of *tcFGE*_Cu_S and *tcFGE*_Cd shows that the transition from trigonal to tetrahedral coordination moves the metal again by 0.8 Å and reduces the distance between the metal and the position of H₂O₁ and H₂O₂ down to 2.8 Å and 2.7 Å (Fig. 3). Tyr273 adopts the same position as in *tcFGE*_Ag, confirming that the orientation of this residue is dependent on the presence of substrate, rather than on the geometry of the metal complex (Fig. 3). Otherwise, the *tcFGE*_Cu_S and *tcFGE*_Cd structures are remarkably similar, suggesting that most active site reorganization that occurs en route from the Cu(i) resting state to the reactive complex may be induced by substrate-binding.

Trp228 and Ser266 form an anion binding site

The structure of *tcFGE*_Cu_S identifies the position of H₂O₁ and H₂O₂ as the only available pocket for O₂ to initiate an interaction with the copper center. The opposite face of the trigonal complex is blocked by the side chains of Trp228 and Arg276. Hence, further scrutiny of the subtle changes that occur at this site in response to substrate binding may provide additional insight as to how this unusual metal center can activate O₂. The apo structures of *hsFGE* (PDB: 1Y1E)²⁹ and *scFGE* (PDB: 2Q17),³⁰ the metalated forms *tcFGE*_Ag,²⁶ *scFGE*_Cu,²⁷ and *tcFGE*_Cu_S all contain one crystallographic water (H₂O₁) in hydrogen-bonding distance to Ser266 (2.6–2.8 Å) and Trp228 (2.9–3.2 Å, Fig. S7†). In the apo structures the hydrogen bonds with Trp228 are likely weak due to a small bond angle (N–H–O < 130°, Table S4†). Nevertheless, several crystal structures of *hsFGE* with halides bound to the approximate position of H₂O₁ indicate that this pocket is endowed with significant affinity for anionic species.^{32,46} Most interestingly, addition of metal and substrate to *tcFGE* increased the bond angle between H₂O₁ and the indole of Trp228 to 148° (*tcFGE*_Ag) and 157° (*tcFGE*_Cu_S) (Table S4†), suggesting that the presence of metal and substrate also increases bonding interactions at the O₂-



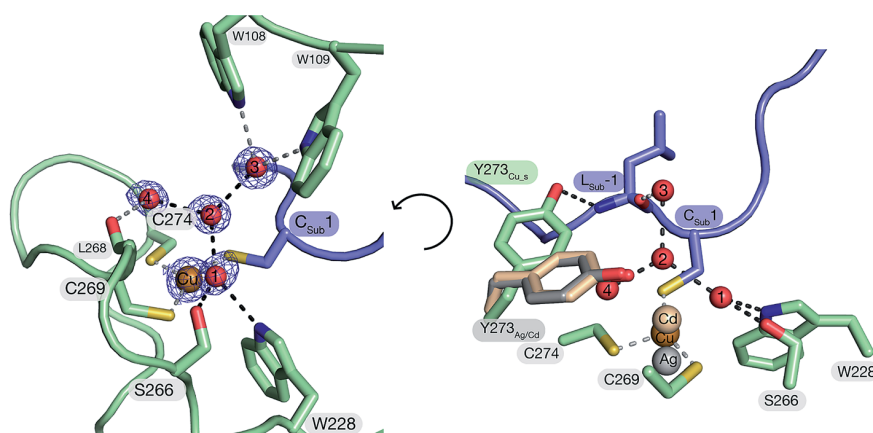


Fig. 3 Left: The primary and secondary coordination sphere of Cu(I) in the crystal structure of *tcFGE_Cu_S* (protein: green, substrate: violet). The $2m|F_o| - D|F_c|$ omit map of Cu(I) and the four crystallographic water molecules (H_2O_1-4) is contoured at σ -level = 1.0. Right: Active site of *tcFGE_Cu_S* including the metals and the side chain of Tyr273 from the structures of *tcFGE_Ag* (gray) and *tcFGE_Cd* (beige).

binding site. Also, based on the observed anion-affinity of this pocket we have proposed previously that H_2O_1 in *tcFGE_Ag* might be a hydroxide that serves as the base that deprotonates the thiol of the incoming peptidyl cysteine.²⁶

Observation of Trp228 by NMR

To examine the hydrogen bonding interaction between H_2O_1 and Trp228 in more detail we used NMR spectroscopy. In addition to directly interrogate the chemical environment of specific protons NMR spectroscopy has the key advantage over X-ray measurements (at low temperatures) that the enzyme is examined in solution and at room temperature. $^1H-^{15}N$ TROSY HSQC spectra (transverse relaxation-optimized spectroscopy heteronuclear single quantum correlation) were measured using uniformly ^{15}N -labelled *tcFGE* and *tcFGE_{W228F}* in apo form and in complex with Cu(I) on a 600 MHz Bruker NMR instrument.⁴⁷ Superposition of these spectra, in combination with a triple resonance HNC0 experiment that distinguishes amide-specific resonances from indole-specific resonances (Fig. S8 and S9[†]),⁴⁸ allowed unambiguous identification of the $H^{\epsilon1}$ and $N^{\epsilon1}$ signals relating to the indole side chain of Trp228 (Fig. S10–S12[†]). The chemical shifts of indole $H^{\epsilon1}$ and $N^{\epsilon1}$ signals are highly sensitive to their environment, reporting on the solvation of the N–H function and the aromatic ring system.^{49,50} Therefore, we reasoned that the Trp228 signal could be used to monitor the protonation state and movements of H_2O_1 . The recorded spectra show that addition of Cu(I) to apo *tcFGE* induced a significant downfield shift of the $H^{\epsilon1}$ signal (Table 1 and Fig. 4). Addition of the substrate to *tcFGE_Cu* induced a further $H^{\epsilon1}$ shift of similar magnitude. Addition of metal also caused slight deshielding of $N^{\epsilon1}$, consistent with increased strength of the hydrogen bond between H_2O_1 and Trp228. In contrast, addition of substrate caused a marked shielding effect. Inspection of the crystal structures shows, that the indole ring of Trp228 is solvent exposed in *tcFGE_Ag* and is buried in *tcFGE_Cu_S* (Fig. S13[†]). Hence, addition of the substrate reduces the polarity of the medium above the indole ring

consistent with the observed shielding effect. The observed deshielding of the indole proton is fully consistent with the proposition that metal and substrate binding increase bonding interactions at the O_2 -binding site.

Complexation of apo *tcFGE* with Ag(I) induced a significantly smaller downfield shift of the $H^{\epsilon1}/N^{\epsilon1}$ signal than observed for Cu(I) (Table 1 and Fig. 4). A likely cause for this small but significant difference is that the bond angle between the sulfur ligands and the metal is larger for Ag(I) than for Cu(I).²⁷ This slight movement of the metal would push H_2O_1 closer to Trp228, and therefore strengthen the interaction between the two. The structures of *tcFGE_Ag* (S–Ag–S: 178°) and *scFGE_Cu* (S–Cu–S: 171°) indeed document a small structural difference.²⁷ Addition of the substrate to *tcFGE_Ag_S* shifted the $H^{\epsilon1}/N^{\epsilon1}$ signal of *tcFGE_Ag_S* to exactly the same position as measured with *tcFGE_Cu_S* (Table 1), suggesting that the two complexes adopt very similar structures. The crystallographic and spectroscopic findings validate Ag(I) as a reliable structural mimic of Cu(I) in FGE.

Using this sensitive NMR methodology, we examined the proposition that H_2O_1 in the *tcFGE_Cu* is stabilized as a hydroxide. Even though H_2O_1 is not a metal ligand, it is possible that the proximity of an additional cationic charge, compounded by the interactions with Ser266 and Trp228 could stabilize a hydroxide at physiological pH. Since the pK_a of water closely matches that of indole rings, a hydrogen bond between indole and hydroxide is likely stronger than between indole and water. Superposition of three $^1H-^{15}N$ TROSY HSQC spectra for *tcFGE_Cu* measured at pH 6, 7 and 8 showed no change of the Trp228 signal relative to other indole-specific resonances (Fig. 4). Based on this observation we concluded that either the pK_a of H_2O_1 is far below 6, or that H_2O_1 is neutral at physiological pH. The second option is far more likely.

Role of Ser266 and Trp266 in the resting state and in catalysis

Finally, we examined the effect of the Ser266 to Ala mutation on the interaction of H_2O_1 with Trp228. $^1H-^{15}N$ TROSY HSQC



Table 1 The observed $H^{\epsilon 1}$ and $N^{\epsilon 1}$ chemical shifts of Trp228 sidechain upon metalation and substrate binding^a

Enzyme	Metal	$\delta H^{\epsilon 1}/N^{\epsilon 1}$ Trp228, ppm			$\Delta \delta H^{\epsilon 1}/N^{\epsilon 1}$ (Trp228), ppm		
		E	E : M	E : M : S	E \rightarrow E : M	E : M \rightarrow E : M : S	E \rightarrow E : M : S
<i>tcFGE</i>	Ag(I)	9.69/128.93	9.88/129.95	10.26/126.81	0.19/1.02	0.38/−3.14	0.57/−2.12
<i>tcFGE</i>	Cu(I)	9.69/128.93	9.98/130.75	10.22/126.66	0.29/1.82	0.24/−4.09	0.53/−2.27
<i>tcFGE</i> _{S266A}	Cu(I)	9.75/129.29	10.04/130.78	10.22/126.77	0.29/1.49	0.18/−4.01	0.47/−2.52

^a The errors of the given values are estimated to be <0.02 ppm.

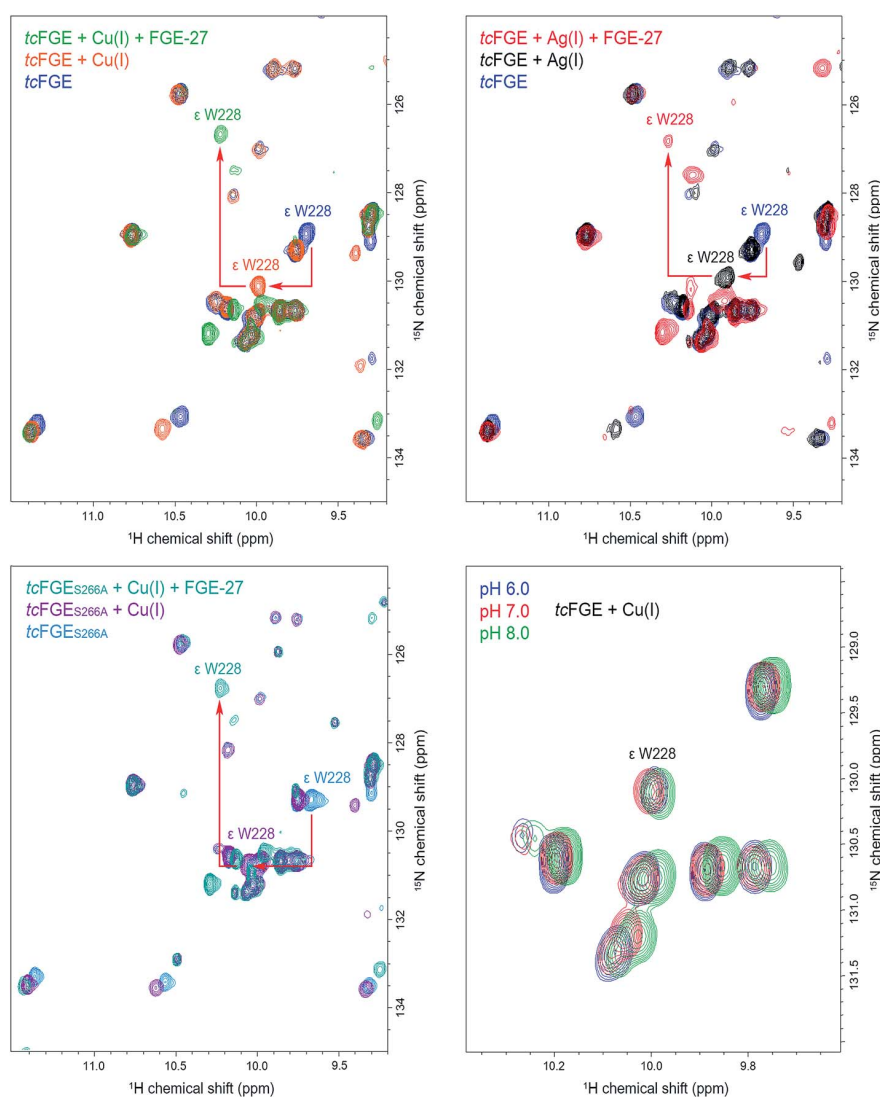


Fig. 4 1H – ^{15}N TROSY HSQC spectra of uniformly ^{15}N -labelled $300 \mu M$ *tcFGE* in 20 mM phosphate buffer, 50 mM NaCl, 2 mM DTT, 25 C. The downfield area of the amide region is presented. Top left: The overlay of the spectra of apo-form *tcFGE*, pH 8.0 (dark blue), *tcFGE* in the presence of $450 \mu M$ $CuSO_4$, pH 8.0 (orange), *tcFGE* in the presence of $450 \mu M$ $CuSO_4$ and 1 mM FGE-27, 2 mM glucose, 2 units (U) glucose oxidase (GO), 200 U catalase (CAT), pH 8.1 (green). Top right: The overlay of the spectra of apo-form *tcFGE*, pH 8.0 (dark blue), *tcFGE* in the presence of $450 \mu M$ $AgNO_3$, pH 8.1 (red), *tcFGE* in the presence of $450 \mu M$ $AgNO_3$ and 1 mM FGE-27, 2 mM glucose, 2 U GO, 200 U CAT, pH 8.1 (red). Bottom left: The overlay of the spectra of apo-form *tcFGE*_{S266A}, pH 8.0 (blue), *tcFGE*_{S266A} in the presence of $450 \mu M$ $CuSO_4$, pH 8.0 (magenta), *tcFGE*_{S266A} in the presence of $450 \mu M$ $CuSO_4$ and 1 mM FGE-27, 2 mM glucose, 2 U GO, 200 U CAT, pH 8.1 (dark cyan). Bottom right: The overlay of the spectra of *tcFGE* in the presence of $450 \mu M$ $CuSO_4$ measured at different pH.



spectra recorded for *tcFGE*_{S266}, *tcFGE*_{S266A}-Cu and *tcFGE*_{S266A}-Cu_S revealed only small or no deshielding effects on the H^{ε1}/N^{ε1} signal, suggesting that the absence of the Ser266 side chain does not affect the interaction between H₂O_1 and Trp228. Hence, the role of Ser266 in immobilizing H₂O_1 in the *tcFGE*-Cu_S complex is limited. In stark contrast, the same mutation causes a 50-fold reduction in catalytic turnover (*k*_{cat}) without affecting substrate binding (*K*_m, Table 2).²⁵ The kinetic and spectroscopic properties of *tcFGE*_{S266A} suggest that Ser266 provides a hydrogen bond that become important after substrate-binding. Similarly, mutation of Trp228 to Phe (*tcFGE*_{W288F}) also specifically affects *k*_{cat} without hurting substrate binding. The lack of this hydrogen bond donor in the O₂-binding site is likely to contribute to the demise of catalysis by this variant.

The unusual resting state of FGE

The active site of FGE is unusual for several reasons. First, catalysts that use a bis-cysteine coordinated Cu(I) as a starting point for O₂-dependent C–H cleavage is unexplored territory. A large number of studies have shown quite conclusively that imidazole, amine, amide and thioether ligands are particularly well suited to facilitate copper-mediated O₂-activation, not the least because these ligands stabilize high-valent copper species.^{1–3,51,52} From this perspective, bis- or tris-thiolate coordination spheres seem odd choices for designing copper-dependent oxidation catalysts. Thiolate ligands stabilize Cu(I) which should render electron transfer to O₂ more difficult. On the other hand, thiolates are intrinsically vulnerable to oxidation which may open efficient paths for catalyst destruction. A specific complication of the FGE-catalyzed reaction is that the two active site cysteines and the substrate contain altogether six chemically equivalent cysteinyl C_β-H bonds. Hence, the geometry of the active site must ensure that the reactive oxygen species exclusively attacks the pro-(*R*)-β-hydrogen on the substrate.¹⁹ The structure of *tcFGE*-Cu_S visualizes how the enzyme achieves regiospecificity. All three cysteines approach Cu(I) in a similar angle (C_β-S-Cu: 103–120°) and place their C_β at a similar distance to the copper center (3.2–3.5 Å). However, only the pro-(*R*)-β-hydrogen of the substrate points towards the presumed O₂-binding site (Fig. 5).

The second unusual feature of FGE is that the substrate and the reactive oxygen species both coordinate to the metal.^{1,27} Metal-dependent oxidases often bind O₂ *via* displacement of a weakly bound metal ligand, such as water, at a pre-formed coordination

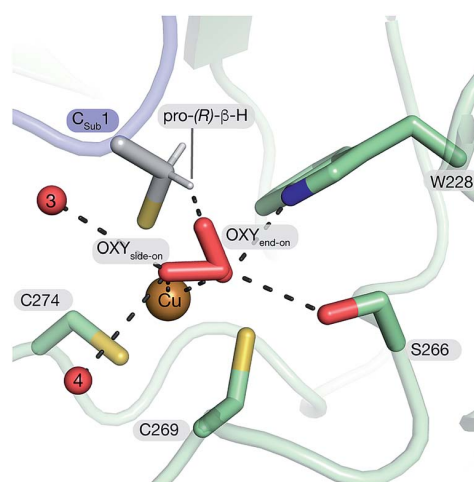


Fig. 5 Structure-based model of *tcFGE*-Cu_S with superoxide coordinated to Cu(II) in side-on mode (η^1) or end-on mode (η^2).

site. Formation of the coordination bond is accompanied by reduction of O₂ to superoxide *via* inner-sphere electron transfer. For example, PHM coordinates Cu(I) *via* the side chains of two histidines and one methionine in a near tetrahedral geometry. The fourth ligand is a water molecule that is displaced by O₂.^{9,10} In LPMO Cu(I) is coordinated by a histidine brace ($N\alpha/N\pi$) and the side chain of a second histidine.⁷ The square planar coordination sphere is completed by a chloride or a water molecule filling the dedicated O₂ binding site. CAO coordinates copper with three histidines and a loosely bound water. Displacement by O₂ occurs without change in the coordination geometry.¹¹ We note that the path leading to the cupric superoxide in CAO is controversial. In the conventional mechanism O₂ forms a coordination bond with Cu(I).⁵³ An alternative proposal suggests that prebound O₂ is first reduced by a protein-derived redox cofactor, and then combines as superoxide with Cu(II).⁵⁴

In contrast to these enzymes, the substrate complex of *tcFGE* contains no preformed coordination site for a fourth ligand. The tris-thiolate Cu(I) complex is completely planar and the closest crystallographic waters (H₂O_1 and H₂O_2) are too distant for attractive interactions. Hence, direct binding of O₂ would require a change in coordination geometry. The structure of *tcFGE* in complex with Cd(II) showed a tetrahedral coordination sphere around this bivalent metal. Therefore it is plausible that oxidation of Cu(I) to Cu(II) may trigger a change from trigonal planar to tetrahedral coordination.²⁶ Importantly, if O₂-coordination requires structural change, and this change requires oxidation of Cu(I), it is unlikely that this redox reaction occurs *via* inner-sphere electron transfer as in other mononuclear copper enzymes. A more plausible scenario is that O₂ first binds to the active site without direct contact to the metal (C, Fig. 6). Outer sphere electron transfer from Cu(I) to pre-bound O₂ followed by coordination would produce the predicted cupric superoxide complex (D). The efficiency of outer sphere electron transfer would critically rely on a binding site that increases the redox potential of pre-bound O₂.

Table 2 The Characterization of the catalytic activity of *tcFGE* variants using HPLC-based kinetic assay^a

Enzyme	<i>k</i> _{cat} , [min ⁻¹]	<i>K</i> _m , [μM]	<i>k</i> _{cat} / <i>K</i> _m , [min ⁻¹ M ⁻¹]
<i>tcFGE</i>	1.4 ± 0.07	530 ± 40	2700 ± 200
<i>tcFGE</i> _{S266A} ^b	0.025 ± 0.002	520 ± 240	49 ± 8
<i>tcFGE</i> _{W228F}	0.0073 ± 0.0002	390 ± 20	19 ± 1

^a Kinetic parameters were determined in the presence of 2 μM CuSO₄, 5 mM DTT, 50 mM NaCl and 50 mM Tris buffer pH 8.0, at 25°C. Rate determined by monitoring the formation of fGly. ^b From ref. 25.



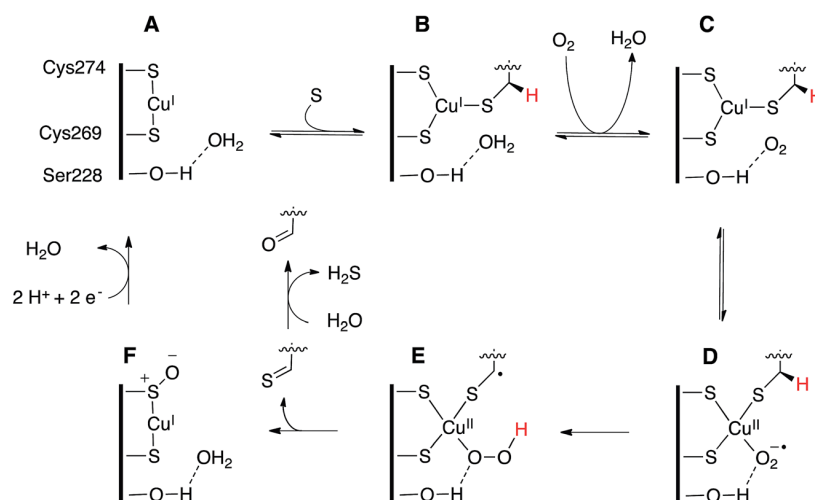


Fig. 6 Proposed catalytic mechanism of FGE. The reaction starts with sequential binding of substrate and O_2 forming first complex B (crystallized) and then C. Outer sphere electron transfer from $Cu(I)$ to O_2 could produce the $Cu(II)$ superoxide species D (modelled in Fig. 5). Homolytic abstraction of the pro-(*R*)- β -hydrogen atom from the substrate (E), followed by release and hydrolysis of the oxidized peptide product and reduction of the oxidized enzyme form (F) return the enzyme to the catalytically active resting state A.

This requirement highlights the third unusual feature of FGE. The presumed O_2 -binding site in *tcFGE* is lined with an extensive array of hydrogen bond donors. Hydrogen-bonding has been recognized as important modulator of the stability and reactivity of copper-coordinated oxygen species in enzymes and in abiotic catalysts.^{1,2,55–57} For example, hydrogen-bonding to the proximal oxygen atom of a synthetic cupric superoxide complex has been shown to increase its stability and its propensity for hydrogen atom abstraction in direct correlation with hydrogen-bonding strength.⁵⁵ Furthermore, specific hydrogen-bonding to the distal or to the proximal oxygen has been shown to affect the activity of copper-oxygen adducts in very different ways.⁵⁶ On the other hand, oxygen-binding pockets lined by four well-positioned hydrogen bond donors are rare among metal-dependent catalysts. Studies on synthetic cages showed that solvation of O_2 with multiple amide- or amine-donated hydrogen bonds can significantly stabilize superoxide and peroxide anions.^{58,59} The acidic O_2 -binding pocket in FGE may play a similar stabilizing role and thereby facilitate electron transfer from the electron-rich tris-thiolate $Cu(I)$ to O_2 . The observation that mutation of Ser266 or Trp228 dramatically reduce the activity of *tcFGE* is consistent with this interpretation.

Conclusions

In this report we describe the crystal structure of *tcFGE* in complex with $Cu(I)$ and a 17-residue peptide. Comparisons of this complex with the structures of *tcFGE* bound to $Ag(I)$ or $Cd(II)$ highlight substrate-induced changes that may prime the active site for O_2 -binding. On the other hand, we also found that substrate binding does not induce a preformed coordination site for O_2 at the metal center. These observations raise the possibility that FGE activates O_2 in a different way than other

metallo-enzymes. Specifically, we propose that O_2 first extracts an electron from $Cu(I)$ via outer sphere electron transfer before combining to a cupric superoxide species. This and other testable hypotheses emanating from our structural investigating set the stage for deciphering the detailed mechanism of this unusual oxidation catalyst. We also hope to inspire the synthesis and characterization of abiotic catalysts that exploit this reactivity of bis- and tris-thiolate coordinated copper species.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We thank Prof. Dr Oliver Einsle (Univ. Freiburg) for the access to the equipment used for anaerobic crystallization, the Swiss Lightsource (Villigen, Switzerland) for access to the beamline facilities and beamline staff for support. This project was supported by a starting grant from the European Research Council (ERC-2013-StG 336559), the NCCR for Molecular Systems Engineering and by the ‘‘Professur f ur Molekulare Bionik’’.

References

- 1 E. I. Solomon, D. E. Heppner, E. M. Johnston, J. W. Ginsbach, J. Cirera, M. Qayyum, M. T. Kieber-Emmons, C. H. Kjaergaard, R. G. Hadt and L. Tian, *Chem. Rev.*, 2014, **114**, 3659–3853.
- 2 C. E. Elwell, N. L. Gagnon, B. D. Neisen, D. Dhar, A. D. Spaeth, G. M. Yee and W. B. Tolman, *Chem. Rev.*, 2017, **117**, 2059–2107.



- 3 W. Keown, J. B. Gary and T. D. Stack, *J. Biol. Chem.*, 2017, **22**, 289–305.
- 4 S. E. Allen, R. R. Walvoord, R. Padilla-Salinas and M. C. Kozlowski, *Chem. Rev.*, 2013, **113**, 6234–6458.
- 5 J. T. Rubino and K. J. Franz, *J. Inorg. Biochem.*, 2012, **107**, 129–143.
- 6 R. J. Quinlan, M. D. Sweeney, L. L. Leggio, H. Otten, J. C. N. Poulsen, K. S. Johansen, K. B. R. M. Krogh, C. I. Jorgensen, M. Tovborg, A. Anthonsen, T. Tryfona, C. P. Walter, P. Dupree, F. Xu, G. J. Davies and P. H. Walton, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 15079–15084.
- 7 K. E. Frandsen, T. J. Simmons, P. Dupree, J. C. Poulsen, G. R. Hemsworth, L. Ciano, E. M. Johnston, M. Tovborg, K. S. Johansen, P. von Freiesleben, L. Marmuse, S. Fort, S. Cottaz, H. Driguez, B. Henrissat, N. Lenfant, F. Tuna, A. Baldansuren, G. J. Davies, L. Lo Leggio and P. H. Walton, *Nat. Chem. Biol.*, 2016, **12**, 298–303.
- 8 L. Cao, O. Calderaru, A. C. Rosenzweig and U. Ryde, *Angew. Chem., Int. Ed.*, 2018, **57**, 162–166.
- 9 S. T. Prigge, B. A. Eipper, R. E. Mains and L. M. MAMzel, *Science*, 2004, **304**, 864–867.
- 10 S. T. Prigge, A. S. Kolhekar, B. A. Eipper, R. E. Mains and M. Amzel, *Nat. Struct. Biol.*, 1999, **6**, 976–983.
- 11 C. M. Wilmot, J. Hajdu, M. J. McPherson, P. F. Knowles and S. E. Phillips, *Science*, 1999, **286**, 1724–1728.
- 12 T. V. Vendelboe, P. Harris, Y. Zhao, T. S. Walter, K. Harlos, K. El Omari and H. E. M. Christensen, *Sci. Adv.*, 2016, **2**, e1500980.
- 13 J. Y. Lee and K. D. Karlin, *Curr. Opin. Chem. Biol.*, 2015, **25**, 184–193.
- 14 A. Changela, K. Chen, Y. Xue, J. Holschen, C. E. Outten, T. V. OHalloran and A. Mondragon, *Science*, 2003, **301**, 1383–1387.
- 15 A. K. Wernimont, D. L. Huffman, A. L. Lamb, T. V. OHalloran and A. C. Rosenzweig, *Nat. Struct. Biol.*, 2000, **7**, 766–771.
- 16 A. K. Boal and A. C. Rosenzweig, *Chem. Rev.*, 2009, **109**, 4760–4779.
- 17 L. Banci, I. Bertini, F. Cantini, I. C. Felli, L. Gonnelli, N. Hadjiladis, R. Pierattelli, A. Rosato and P. Voulgaris, *Nat. Chem. Biol.*, 2006, **2**, 367–368.
- 18 P. G. Holder, L. C. Jones, P. M. Drake, R. M. Barfield, S. Banas, G. W. de Hart, J. Baker and D. Rabuka, *J. Biol. Chem.*, 2015, **290**, 15730–15745.
- 19 M. Knop, P. Engi, R. Lemnar and F. P. Seebeck, *ChemBioChem*, 2015, **16**, 2147–2150.
- 20 T. Dierks, B. Schmidt, L. V. Borissenko, J. H. Peng, A. Preusser, M. Mariappan and K. von Figura, *Cell*, 2003, **113**, 435–444.
- 21 M. P. Cosma, S. Pepe, I. Annunziata, R. F. Newbold, M. Grompe, G. Parenti and A. Ballabio, *Cell*, 2003, **113**, 445–456.
- 22 I. S. Carrico, B. L. Carlson and C. R. Bertozzi, *Nat. Chem. Biol.*, 2007, **3**, 321–322.
- 23 M. J. Appel and C. R. Bertozzi, *ACS Chem. Biol.*, 2015, **10**, 72–84.
- 24 T. Krüger, T. Dierks and N. Sewald, *Biol. Chem.*, 2019, **4000**, 289–297.
- 25 M. Knop, T. Q. Dang, G. Jeschke and F. P. Seebeck, *ChemBioChem*, 2017, **18**, 161–165.
- 26 M. Meury, M. Knop and F. P. Seebeck, *Angew. Chem., Int. Ed.*, 2017, **56**, 8115–8119.
- 27 M. J. Appel, K. K. Meier, J. Lafrance-Vanasse, H. Lim, C. L. Tsai, B. Hedman, K. O. Hodgson, J. A. Tainer, E. I. Solomon and C. R. Bertozzi, *Proc. Natl. Acad. Sci. U. S. A.*, 2019, **116**(12), 5370–5375.
- 28 M. Jaskolski, M. Gilski, Z. Dauter and A. Wlodawer, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2007, **63**, 611–620.
- 29 T. Dierks, A. Dickmanns, A. Preusser-Kunze, B. Schmidt, M. Mariappan, K. von Figura, R. Ficner and M. G. Rudolph, *Cell*, 2005, **121**, 541–552.
- 30 B. L. Carlson, E. R. Ballister, E. Skordalakes, D. S. King, M. A. Breidenbach, S. A. Gilmore, J. M. Berger and C. R. Bertozzi, *J. Biol. Chem.*, 2008, **283**, 20117–20125.
- 31 A. J. Doig, M. W. MacArthur, B. J. Stapley and J. M. Thornton, *Protein Sci.*, 1997, **6**, 147–155.
- 32 D. Roeser, A. Preusser-Kunze, B. Schmidt, K. Gasow, J. G. Wittmann, T. Dierks, K. von Figura and M. G. Rudolph, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 81–86.
- 33 M. Knop, R. Lemnar and F. P. Seebeck, *ChemBioChem*, 2017, **18**, 1755–1761.
- 34 E. L. Smith, J. P. Giddens, A. T. Iavarone, K. Godula, L. X. Wang and C. R. Bertozzi, *Bioconjugate Chem.*, 2014, **25**, 788–795.
- 35 J. H. Peng, S. Alam, K. Radhakrishnan, M. Mariappan, M. G. Rudolph, C. May, T. Dierks, K. von Figura and B. Schmidt, *FEBS J.*, 2015, **282**, 3262–3274.
- 36 T. Krüger, S. Weiland, G. Falck, M. Gerlach, M. Boschanski, S. Alam, K. M. Müller, T. Dierks and N. Sewald, *Angew. Chem., Int. Ed.*, 2018, **57**, 7245–7249.
- 37 T. Urbic and K. A. Dill, *J. Am. Chem. Soc.*, 2018, **140**, 17106–17113.
- 38 A. Biela, N. N. Nasief, M. Betz, A. Heine, D. Hangauer and G. Klebe, *Angew. Chem., Int. Ed.*, 2013, **52**, 1822–1828.
- 39 M. C. Chervenak and E. J. Toone, *J. Am. Chem. Soc.*, 1994, **116**, 10533–10539.
- 40 P. W. Snyder, J. Mecinovic, D. T. Moustakas, S. W. r. Thomas, M. Harder, E. T. Mack, M. R. Lockett, A. Héroux, W. Sherman and G. M. Whitesides, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 17889–17894.
- 41 F. Biedermann, W. M. Nau and H. J. Schneider, *Angew. Chem., Int. Ed.*, 2014, **53**, 11158–11171.
- 42 P. C. Roehm and J. M. Berg, *J. Am. Chem. Soc.*, 1998, **120**, 13083–13087.
- 43 S. J. Chiou, C. G. Riordan and A. L. Rheingold, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **1000**, 3695–3700.
- 44 A. Dey, T. A. Okamura, N. Ueyama, B. Hedman, K. O. Hodgson and E. I. Solomon, *J. Am. Chem. Soc.*, 2005, **127**, 12046–12053.
- 45 X. Yang, S. Niu, T. Ichiye and L. S. Wang, *J. Am. Chem. Soc.*, 2004, **126**, 15790–15794.
- 46 D. Roeser, B. Schmidt, A. Preusser-Kunze and M. G. Rudolph, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2007, **63**, 621–627.



- 47 K. Pervushin, R. Riek, G. Wider and K. Wüthrich, *Proc. Natl. Acad. Sci. U. S. A.*, 1997, **94**, 12366–12371.
- 48 S. Grzesiek and A. Bax, *J. Magn. Reson.*, 1992, **96**, 432–440.
- 49 W. Sicinska, W. M. Westler and H. F. DeLuca, *Proteins*, 2005, **61**, 461–467.
- 50 E. T. Molloy, D. E. Metzler, A. Kintanar, H. Kagamiyama, H. Hayashi, K. Hirotsu and I. Miyahara, *Biochemistry*, 1997, **36**, 615–625.
- 51 S. M. Adam, G. B. Wijeratne, P. J. Rogler, D. E. Diaz, D. A. Quist, J. J. Liu and K. D. Karlin, *Chem. Rev.*, 2018, **118**, 10840–11022.
- 52 L. Ciano, G. J. Davies, W. B. Tolman and P. H. Walton, *Nat. Catal.*, 2018, **1**, 571–577.
- 53 E. M. Shepard, K. M. Okonski and D. M. Dooley, *Biochemistry*, 2008, **47**, 13907–13920.
- 54 S. A. Mills, Y. Goto, Q. Su, J. Plastino and J. P. Klinman, *Biochemistry*, 2002, **41**, 10577–10584.
- 55 M. Bhadra, J. Y. C. Lee, R. E. Cowley, S. Kim, M. A. Siegler, E. I. Solomon and K. D. Karlin, *J. Am. Chem. Soc.*, 2018, **140**, 9042–9045.
- 56 S. I. Mann, T. Heinisch, T. R. Ward and A. S. Borovik, *J. Am. Chem. Soc.*, 2017, **139**, 17289–17292.
- 57 S. Hong, Y.-M. Lee, K. Ray and W. Nam, *Coord. Chem. Rev.*, 2017, **334**, 25–42.
- 58 N. Lopez, D. J. Graham, R. J. McGuire, G. E. Alliger, Y. Shao-Horn, C. C. Cummins and D. G. Nocera, *Science*, 2012, **335**, 450–453.
- 59 E. W. Dahl, J. J. Kiernicki, M. Zeller and N. K. Szymczak, *J. Am. Chem. Soc.*, 2018, **140**, 10075–10079.



Structure of formylglycine-generating enzyme in complex with copper and substrate reveals a pocket for binding and activation of molecular oxygen

Dzmitry A. Miarzlou¹, Florian Leisinger¹, Daniel Joss¹, Daniel Häussinger¹, and Florian P. Seebeck¹

¹ Department of Chemistry, University of Basel, Mattenstrasse 24a, Basel 4002, Switzerland

SUPPORTING INFORMATION

SUPPLEMENTARY FIGURES

Table S1 – Oligonucleotides used for site directed mutagenesis of *tcFGE* variants.

W228F_S	5'-TGGCGGGTAACGTTTTTGAATGGT-3'
W228F_AS	5'-ACCATTCAAAAACGTTACCCGCCA-3'
S266A_S	5'-CAAGGGTGGCGCCTTCCTGT-3'
S266A_AS	5'-ACAGGAAGGCCACCCTTG-3'

Sequence of *tcFGE*

GHHHHHHAENLYFQGHMPSFDFDI PRRSPQEIAKGMVAIPGGTFRMGGEDPDAFPEDGEGPVRTVRLSPFLIDRYAVSNRQFA
AFVKATGYVTDAERYGWSFVFHAAHVAPGTPVMDAVVPEAPWWWAVPGAYWKAPEGPGSSITDRPNHPVVHVSWNDAVAYATWA
GKRLPTEAEWEMAARGGLDQARYPWGNELTPRGRHRCNIWQGTFFVHDTGEDGYTGAPVNAFAPNGYGLYNVAGNVWEWCAD
WWSADWHATE SPATRIDPRGPETGTARVTKGGSFLCHESYCNRYRVAARTCNTPDSSAAHTGFRCAADPL

Sequence of *tcFGE*_{W228F}

GHHHHHHAENLYFQGHMPSFDFDI PRRSPQEIAKGMVAIPGGTFRMGGEDPDAFPEDGEGPVRTVRLSPFLIDRYAVSNRQFA
AFVKATGYVTDAERYGWSFVFHAAHVAPGTPVMDAVVPEAPWWWAVPGAYWKAPEGPGSSITDRPNHPVVHVSWNDAVAYATWA
GKRLPTEAEWEMAARGGLDQARYPWGNELTPRGRHRCNIWQGTFFVHDTGEDGYTGAPVNAFAPNGYGLYNVAGNVFEWCAD
WWSADWHATE SPATRIDPRGPETGTARVTKGGSFLCHESYCNRYRVAARTCNTPDSSAAHTGFRCAADPL

Sequence of *tcFGE*_{S266A}

GHHHHHHAENLYFQGHMPSFDFDI PRRSPQEIAKGMVAIPGGTFRMGGEDPDAFPEDGEGPVRTVRLSPFLIDRYAVSNRQFA
AFVKATGYVTDAERYGWSFVFHAAHVAPGTPVMDAVVPEAPWWWAVPGAYWKAPEGPGSSITDRPNHPVVHVSWNDAVAYATWA
GKRLPTEAEWEMAARGGLDQARYPWGNELTPRGRHRCNIWQGTFFVHDTGEDGYTGAPVNAFAPNGYGLYNVAGNVWEWCAD
WWSADWHATE SPATRIDPRGPETGTARVTKGGAFLCHESYCNRYRVAARTCNTPDSSAAHTGFRCAADPL

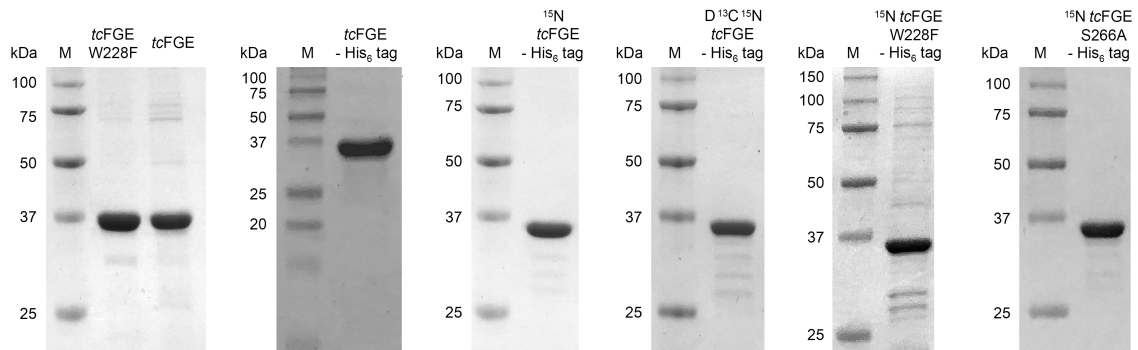


Figure S1 – Characterization of *tcFGE* variants by SDS PAGE.

Table S2 – HR-MS-ESI characterization of *tcFGE* variants and isotopologs.^[a]

Enzyme	Sample description	M _w , [Da]			Labeling efficiency, %
		cal., non-labeled	cal., labeled	observed	
<i>tcFGE</i>		35216.0	-	35215.1	-
<i>tcFGE</i> _{W228F}		35177.0	-	35175.7	-
<i>tcFGE</i>	without His ₆ tag	33470.2	-	33468.9	-
<i>tcFGE</i>	without His ₆ tag	33470.2	33890.9 ^[b]	33889.8	99.7
<i>tcFGE</i> _{W228F}	without His ₆ tag	33431.1	33850.9 ^[b]	33847.3	99.1
<i>tcFGE</i> _{S266A}	without His ₆ tag	33454.2	33874.9 ^[b]	33871.9	99.3
<i>tcFGE</i>	without His ₆ tag	33470.2	37035.2 ^[c]	37012.6	99.4

^[a] The labeling efficiency was calculated as a ratio between the obtained increment $\Delta M_{w \text{ obs.}}$ ($M_{w \text{ obs.}}$ (labeled protein) - $M_{w \text{ cal.}}$ (non-labeled protein)) and the expected increment $\Delta M_{w \text{ cal.}}$ ($M_{w \text{ cal.}}$ (labeled protein) - $M_{w \text{ cal.}}$ (non-labeled protein)).

^[b] Molecular weight calculated with an assumption, that 99% of N in protein replaced with ¹⁵N.

^[c] Molecular weight calculated with an assumption, that 99% of N and C in protein replaced with ¹⁵N and ¹³C respectively, and 99% none water exchangeable ¹H with ²H.

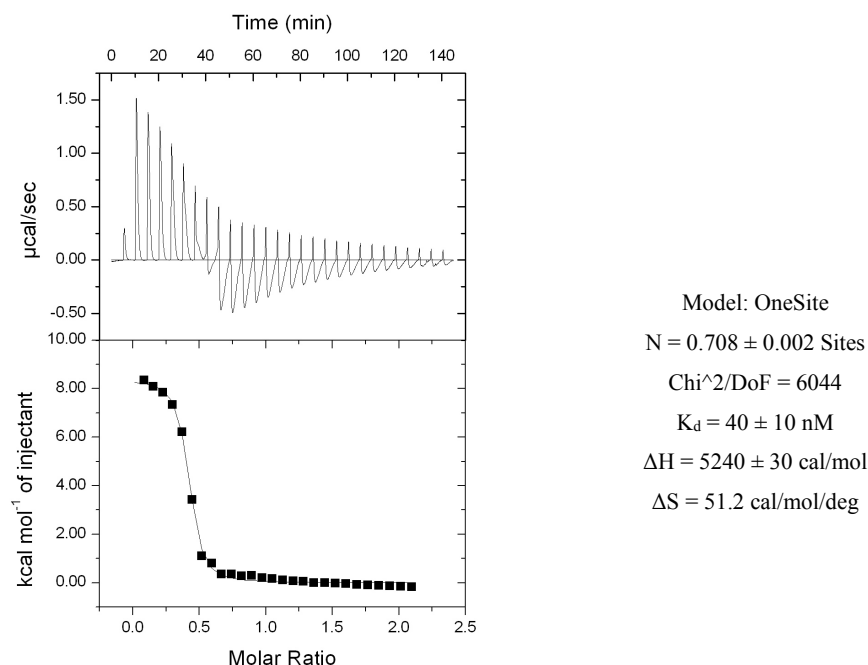


Figure S2 - Isothermal titration calorimetry (ITC). A solution containing 100 μM of *tcFGE*_{W228F} in 50 mM phosphate buffer pH 8.0, 100 mM NaCl and 600 μM DTT was transferred into the sample cell (cell volume 1.42 mL) of an isothermal titration calorimeter (VP200-ITC system, MicroCal LLC). The 1 mM solution of FGE-27, dissolved in the same buffer, was added to the protein solution by syringe in 1 x 2 μL injection followed by 10 μL injections every five minutes. All ITC measurements were conducted at 25°C. Data were plotted as the power needed to maintain the reference and sample cell at the same temperature against time and as kcal/mol of injectant against the molar ratio of ligand and protein. The program Origin7 (OriginLab Corporation) was used to analyze the data.

Table S3 - Data processing and refinement statistics

Structure	Reaction complex with substrate and Cu ⁺
Wavelength (Å)	1.0000
Space group	P 2 ₁ 2 ₁ 2 ₁
Cell dimensions	
a, b, c (Å)	58.4, 71.9, 76.7
α, β, γ (°)	90.0, 90.0, 90.0
Resolution (Å) ^a	46.5 – 1.04 (1.04 – 1.06)
R _{merge} ⁺ (%) ^a	0.054 (0.624)
R _{meas} ⁺⁺ (%) ^a	0.058 (0.695)
R _{pim} ⁺⁺⁺ (%) ^a	0.023 (0.301)
CC ½ ^a	0.999 (0.909)
<I/σ(I)> ^a	19.5 (3.5)
Multiplicity ^a	12.1 (10.1)
Completeness (%) ^a	100 (100)
Refinement statistics	
Resolution (Å)	1.04
No. of reflections	155126
R-work * (%)	16.5
R-free ** (%)	17.5
No. atoms	3825
Protein	3401
Ligand/ion	16
Water	408
B-factors (Å ²)	20.2
Protein	19.2
Ligand/ion	30.1
Water	28.0
r.m.s deviations	
Bond lengths (Å)	0.003
Bond angles (°)	0.59
Ramachandran statistics ***	
(%)	
Favored	96.68
Allowed	3.32
Outliers	0.00

Numbers in parenthesis belong to outer shell.

⁺R_{merge} = $\frac{\sum_{hkl} \sum_i (I_i(hkl) - \langle I(hkl) \rangle)}{\sum_{hkl} \sum_i I_i(hkl)}$ where $I_i(hkl)$ is the observed intensity for a reflection and $\langle I(hkl) \rangle$ is the average intensity obtained from multiple observations of symmetry-related reflections.

⁺⁺R_{meas} = $\frac{\sum_{hkl} [N/(N-1)]^{1/2} \sum_i I_i(hkl) - \langle I(hkl) \rangle}{\sum_{hkl} \sum_i I_i(hkl)}$, where $I_i(hkl)$ is the observed intensity for a reflection, $\langle I(hkl) \rangle$ is the average intensity obtained from multiple observations of symmetry-related reflections and N is the number of observations of intensity $\langle I(hkl) \rangle$.

*R-work = $\frac{\sum |ghkl| |F_{obs}| - |F_{calc}|}{\sum |ghkl| |F_{obs}|}$

**R-free is the R value calculated for 5 % of the data set that was not included into the refinement.

***Molprobit

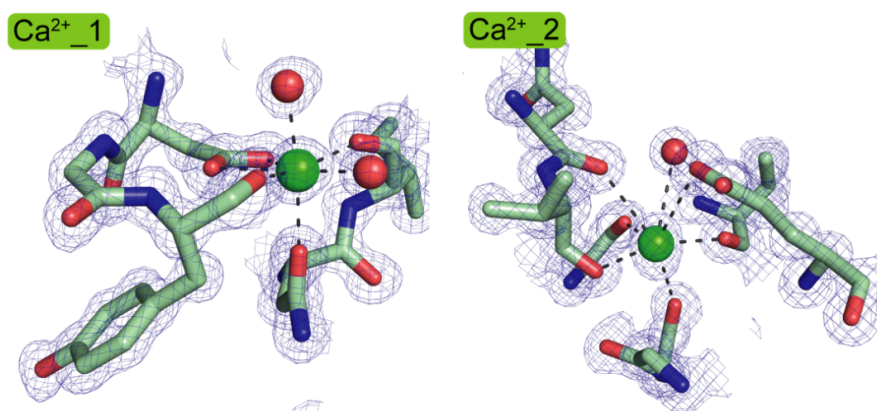


Figure S3 - Calcium bound to conserved calcium-binding site. $2m|F_o|-D|F_c|$ omit map contoured at σ -level = 1.0.

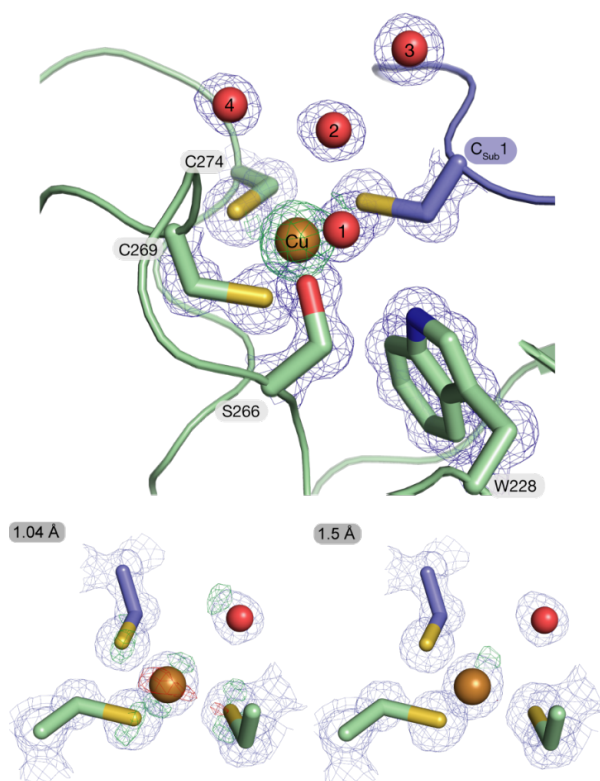


Figure S4 – Top: Detailed view into the active site of *tcFGE_Cu_S* with key active site residues displayed in stick representation with electron density ($2m|F_o|-D|F_c|$) omit map contoured at σ -level = 1.0) in mesh representation (blue). **Bottom:** Additional electron density was found in the active site center as displayed in mesh representation (green, ($m|F_o|-D|F_c|$) map contoured at σ -level = 3.0) which could have been modeled by a Cu (I) ion. After several rounds of refinement two positive and negative electron density peaks of same size came up in the $m|F_o|-D|F_c|$ map. Refinement against the same data scaled to lower resolution of 1.5 Å produced a good fit.

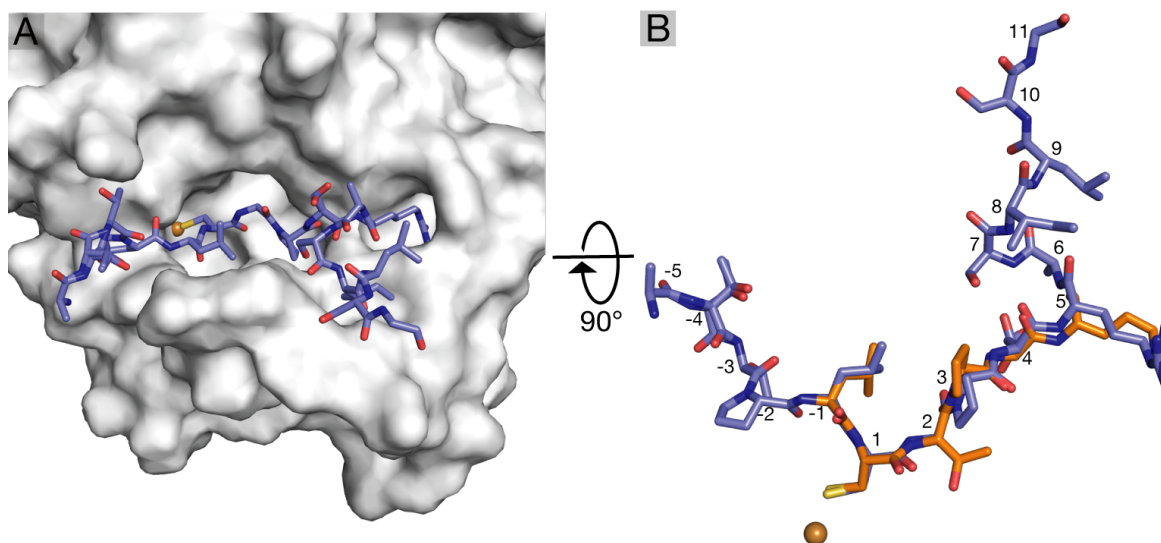


Figure S5 – **A:** Structure of *tcFGE_Cu* (white surface) in complex with substrate (violet, Abz-ATPLCGPSRASILSGR-NH₂). **B:** Superposition of the substrate that was crystallized in complex with *hsFGE* (orange, LCTPSRA, PDB: 2AIK) shows that the cores of both substrates adopt similar conformations.

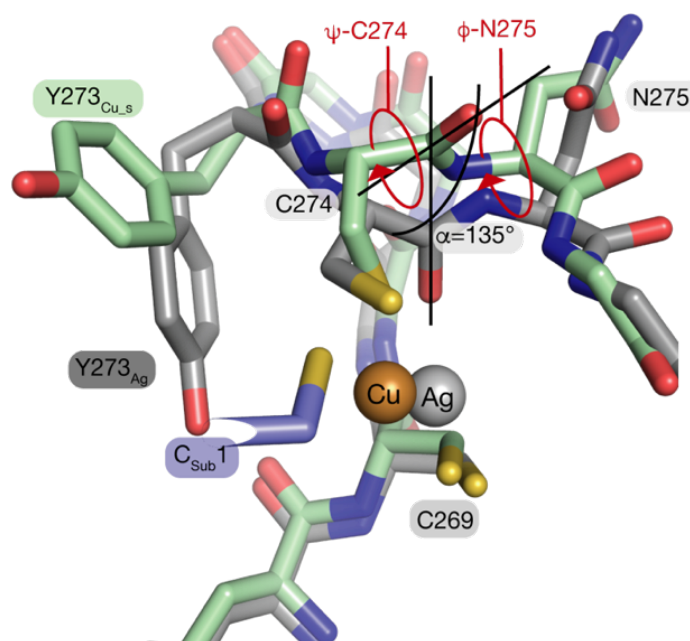


Figure S6 – Rotation of the backbone amide of Cys274 in response to substrate binding. Comparison of the structures from *tcFGE_Ag* and *tcFGE_Cu_S* shows that the plane of this amid bond rotates by 135°. The same rotation was observed in the structure of *tcFGE_Cd*, suggesting that this change is induced by the expansion of the coordination sphere of the metal.

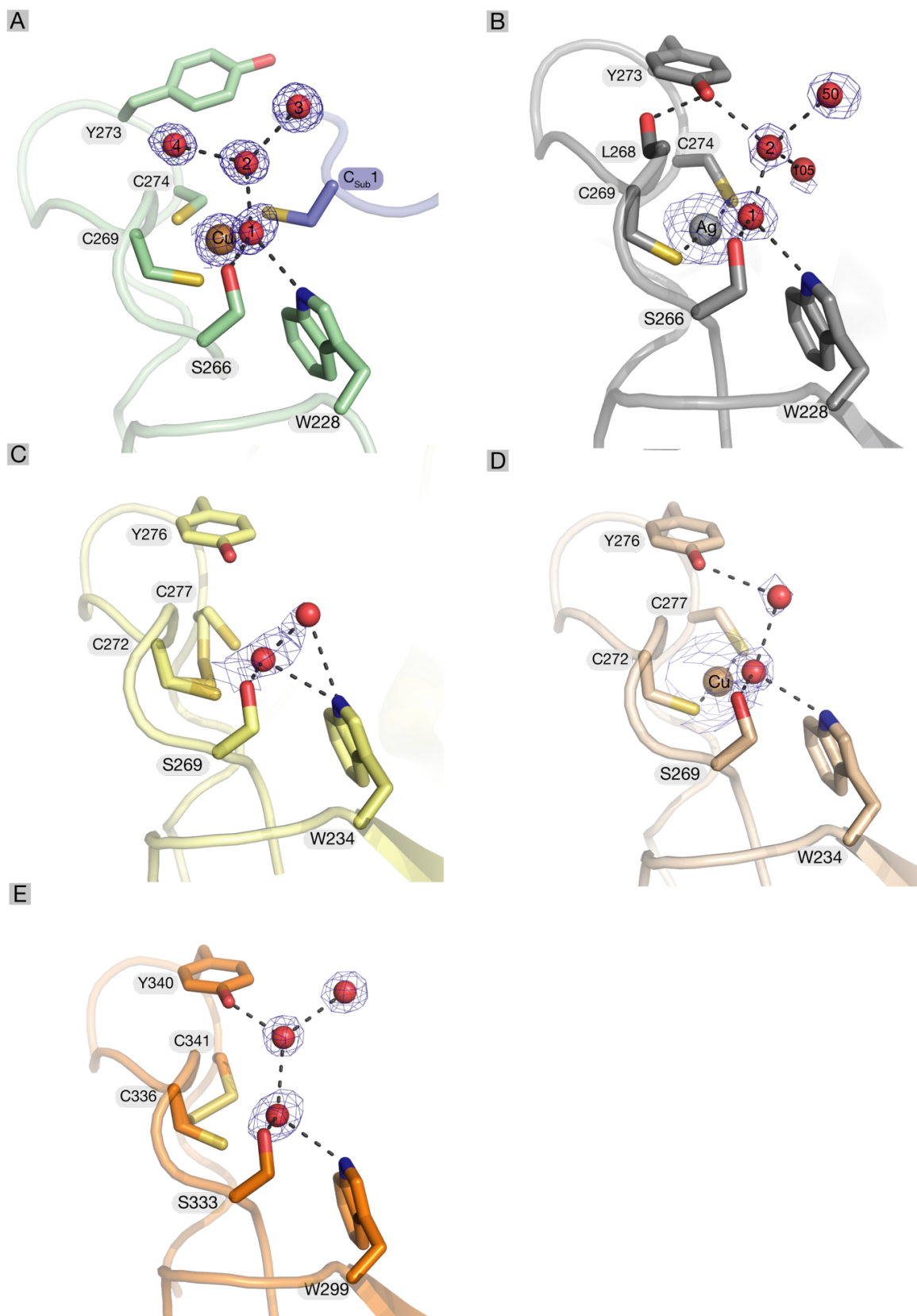


Figure S7 – FGE active sites with crystallographic waters: **A)** *tcFGE_Cu_S* (PDB: XXXX); **B)** *tcFGE_Ag* (PDB: 5NLX); **C)** *scFGE* (PDB: 2Q17); **D)** *scFGE_Cu* (PDB: 6MUJ); **E)** *hsFGE* (PDB: 1Y1E). $2m|F_o|-D|F_c|$ omit maps of the water and metal was contoured at σ -level = 1.1.

Table S4 – The geometry of interactions of conserved active site W and H2O_1.

Structure		Distance	Angle
		O(H ₂ O) – Nε1(W), Å	O(H ₂ O)-Hε1(W)- Nε1(W)
<i>S. coelinicolor</i>	wt (PDB: 2Q17)	3.1	128.0 ⁰
	wt + Cu (PDB: 6MUJ)	3.1	144.1 ⁰
<i>H. sapiens</i>	wt (PDB: 1Y1E)	3.2	127.6 ⁰
<i>T. curvata</i>	wt + Ag(I) (PDB: 5NLX)	2.9	147.5 ⁰
	wt + Cu(I) + S (PDB: XXXX)	2.9	156.7 ⁰

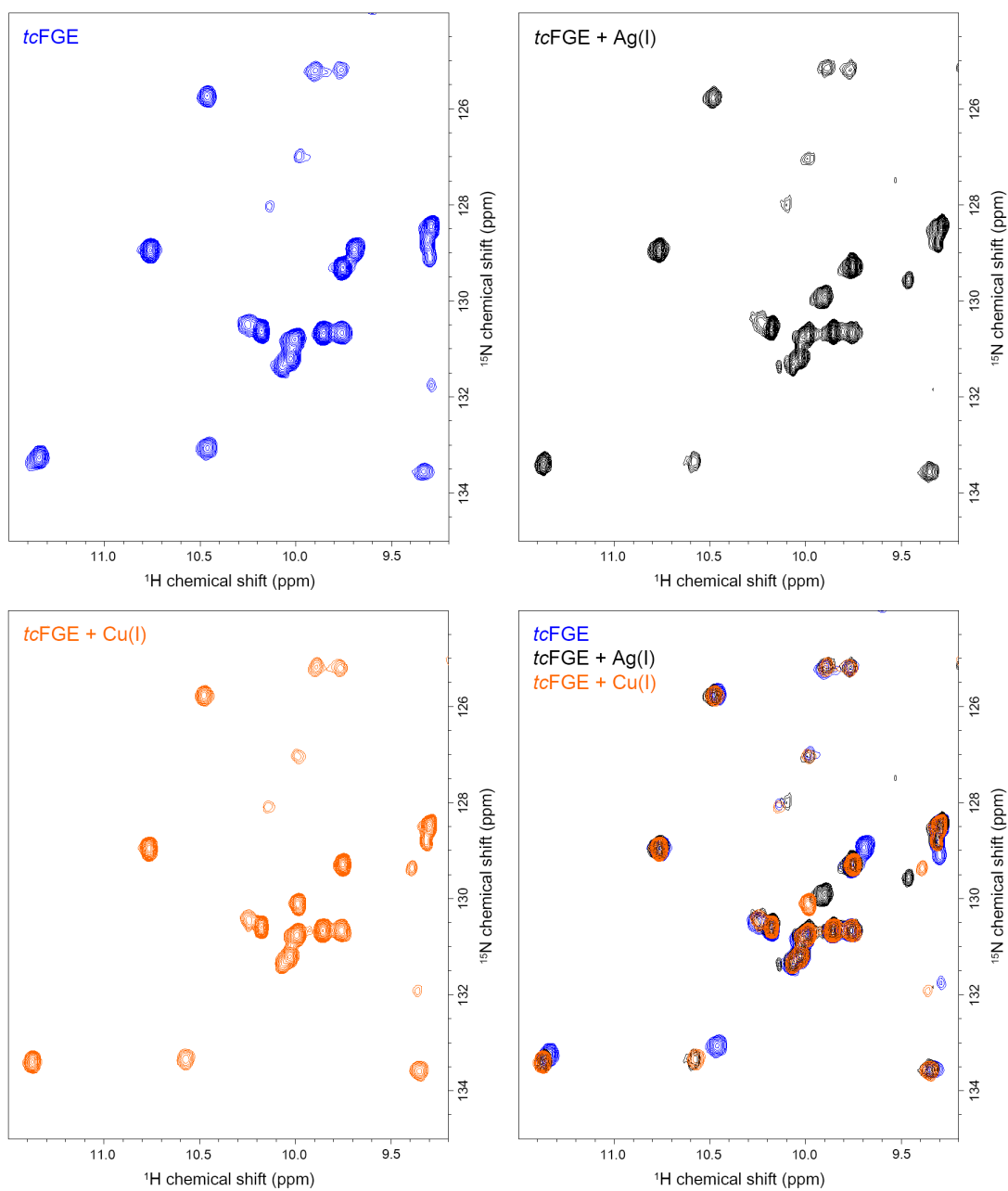


Figure S8 - ^1H - ^{15}N TROSY HSQC spectra of uniformly ^{15}N -labelled *tcFGE* in 20 mM phosphate buffer, 50 mM NaCl, 2 mM DTT, pH 8.0, 25°C. The binding either Ag(I) or Cu(I) in the active site *tcFGE* causes pronounced shifts in the downfield area of the amide region. **Top left:** HSQC spectrum of apo-form *tcFGE* 300 μM . **Top right:** HSQC spectrum of *tcFGE* 300 μM in the presence of 450 μM AgNO_3 . **Bottom left:** HSQC spectrum of *tcFGE* 300 μM in the presence of 450 μM CuSO_4 . **Bottom right:** Overlay of the spectra.

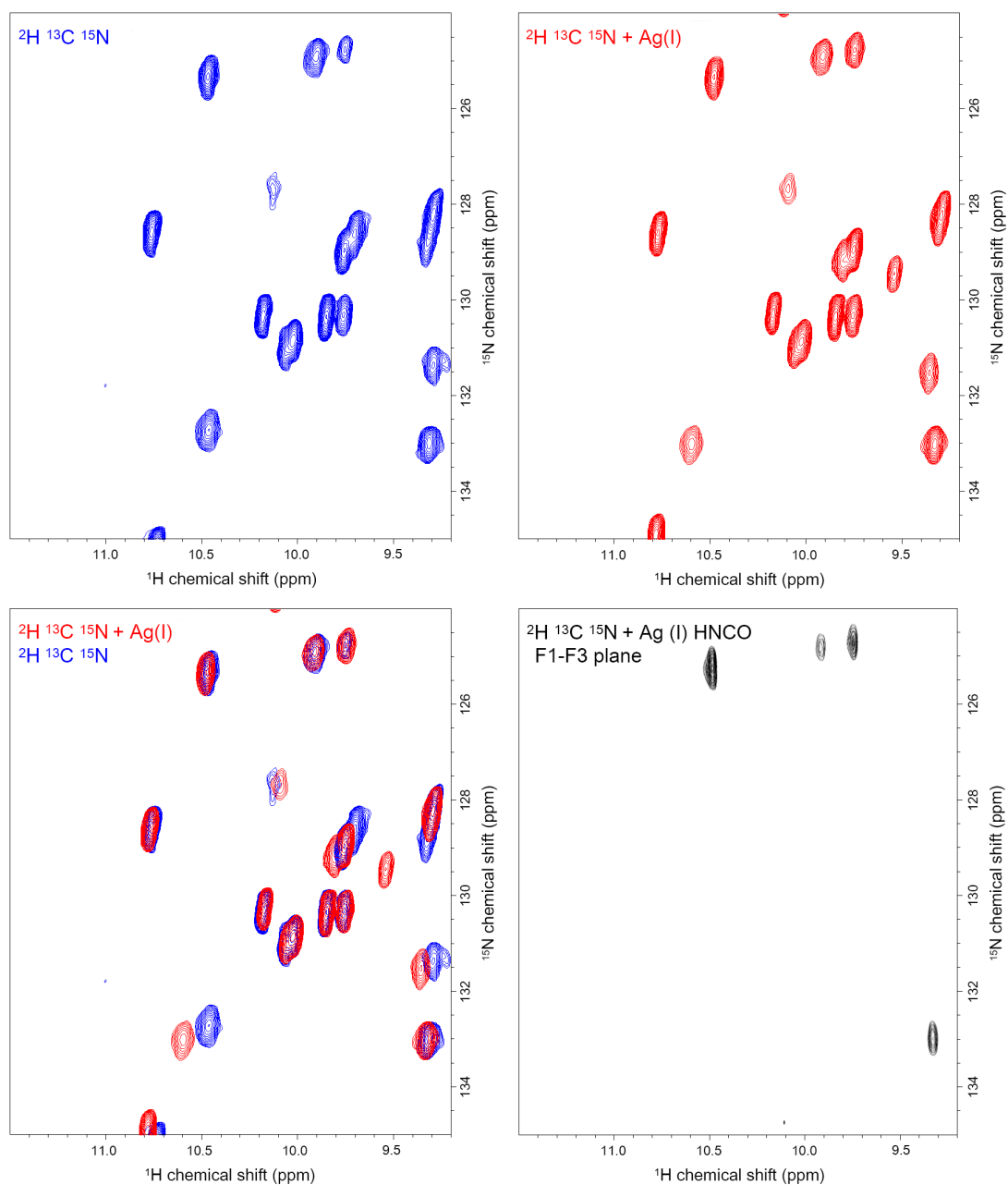


Figure S9 – NMR spectra of ^2H ^{13}C ^{15}N -labelled *tcFGE* in 20 mM phosphate buffer, 75 mM NaCl, 2 mM DTT, pH 8.0, 25°C. The downfield area of the amide region is presented. The comparison of ^1H - ^{15}N TROSY HSQC and F1-F3 (^1H - ^{15}N) plane from TROSY based HNCO experiment for ^2H ^{13}C ^{15}N -labelled *tcFGE* in the presence of 1.5 eq. Ag (I) demonstrated, that significant part of the cross-peaks belongs to $^1\text{H}\epsilon_1$ - $^{15}\text{N}\epsilon_1$ of tryptophan.²⁻³ **Top left:** ^1H - ^{15}N TROSY HSQC spectrum of apo-form ^2H ^{13}C ^{15}N -labelled *tcFGE* 860 μM . **Top right:** ^1H - ^{15}N TROSY HSQC spectrum of ^2H ^{13}C ^{15}N -labelled *tcFGE* 860 μM in the presence of 1.29 mM AgNO_3 . **Bottom left:** The overlay of the spectra of apo- and Ag (I) preloaded ^2H ^{13}C ^{15}N -labelled *tcFGE*. **Bottom right:** F1-F3 plane extracted from the HNCO spectrum of ^2H ^{13}C ^{15}N -labelled *tcFGE* 860 μM in the presence of 1.29 mM AgNO_3 .

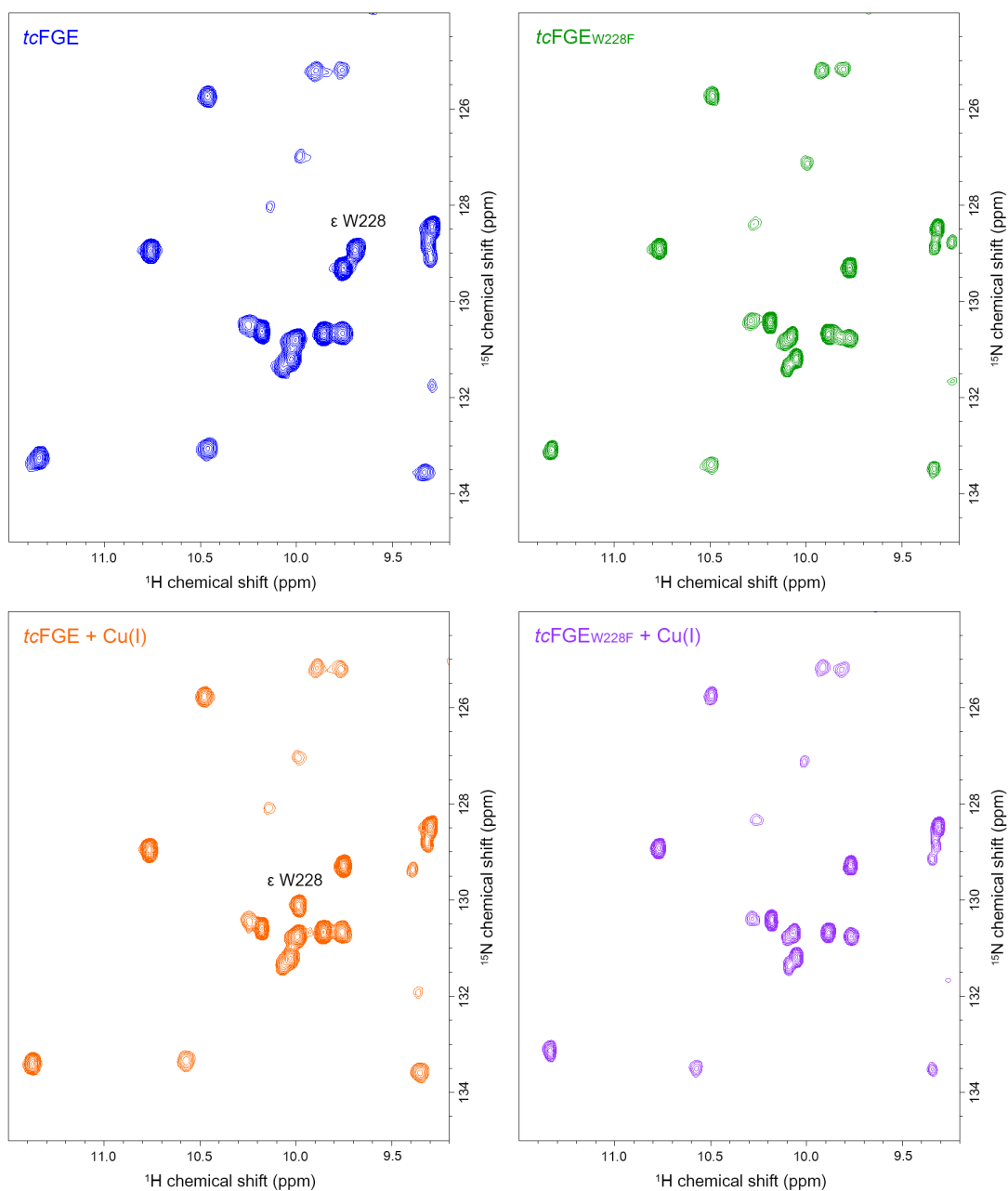


Figure S10 – ^1H - ^{15}N TROSY HSQC spectra of uniformly ^{15}N -labelled *tcFGE* variants in 20 mM phosphate buffer, 50 mM NaCl, 2 mM DTT, pH 8.0, 25°C. The downfield area of the amide region is presented. The mutation W228F allowed to assign the cross peak, which corresponds to $^1\text{H}\epsilon_1$ - $^{15}\text{N}\epsilon_1$ of W228 side chain. **Top left:** HSQC spectrum of apo-form ^{15}N -labelled *tcFGE* 300 μM . **Top right:** HSQC spectrum of apo-form ^{15}N -labelled *tcFGE*_{W228F} 300 μM . **Bottom left:** HSQC spectrum of ^{15}N -labelled *tcFGE* 300 μM in the presence of 450 μM CuSO_4 . **Bottom right:** HSQC spectrum of ^{15}N -labelled *tcFGE*_{W228F} 300 μM in the presence of 450 μM CuSO_4 .

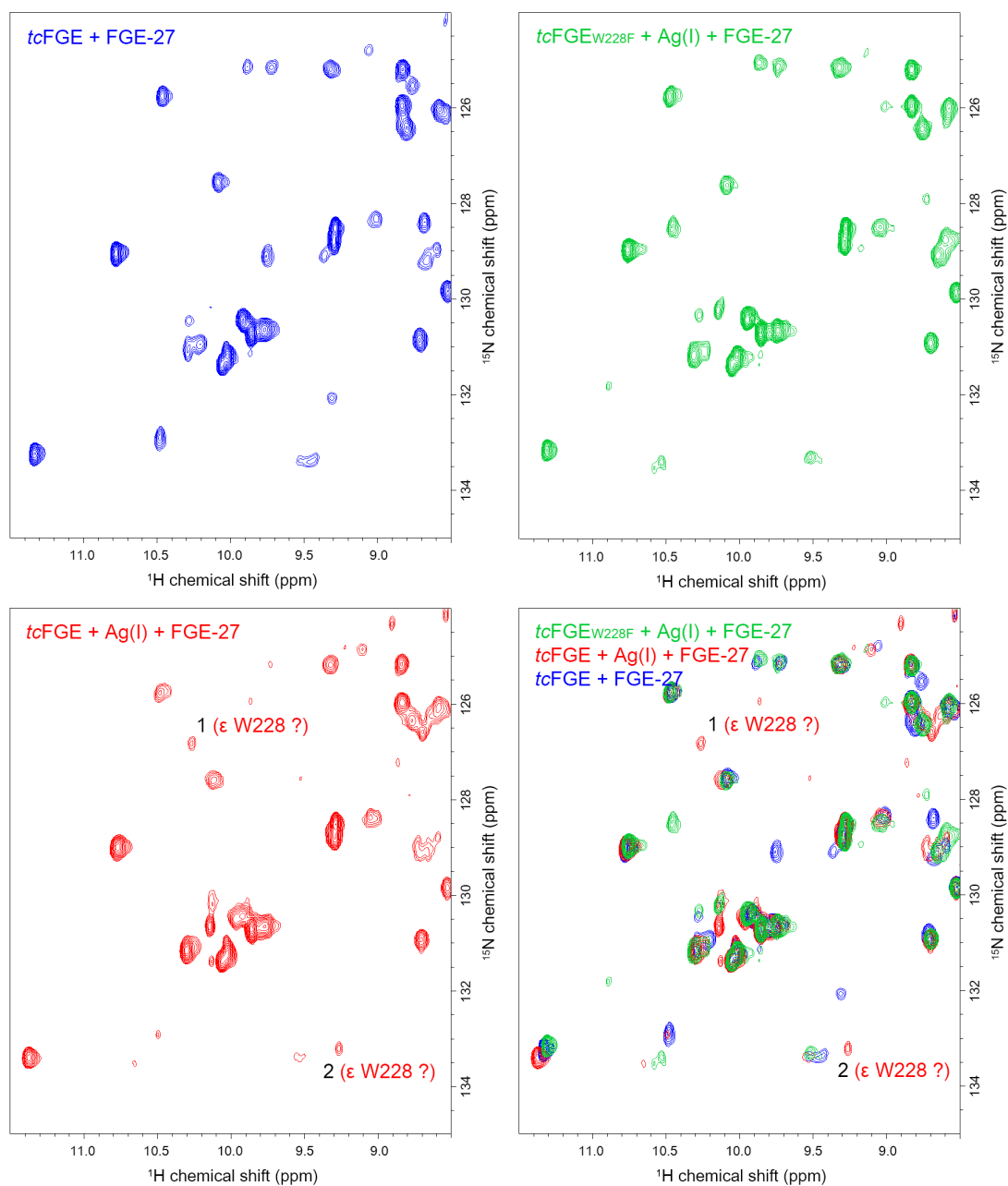


Figure S11 – ^1H - ^{15}N TROSY HSQC spectra of uniformly ^{15}N -labelled *tcFGE* in 20 mM phosphate buffer, 50 mM NaCl, 2 mM DTT, 2 mM Glc, 2 U GO, 200 U CAT, pH 8.1, 25°C. The downfield area of the amide region is presented. Comparing spectra of three systems *tcFGE* : FGE-27, *tcFGE*_{W228F} : Ag(I) : FGE-27 and *tcFGE* : Ag(I) : FGE-27 provides two new cross-peaks 1 (10.26/126.81) and 2 (9.26/133.20), which only appear upon substrate binding to *tcFGE* : Ag (I) complex. These new signals are assigned as possible candidates for the cross-peak $^1\text{H}\epsilon 1$ - $^{15}\text{N}\epsilon 1$ W228, which shifts upon substrate binding. **Tom left:** HSQC spectrum of apo-form *tcFGE* 300 μM in the presence of 1 mM FGE-27. **Top right:** HSQC spectrum of *tcFGE*_{W228F} 300 μM in the presence of 450 μM AgNO₃ and 1 mM FGE-27. **Bottom left:** HSQC spectrum of *tcFGE* 300 μM in the presence of 450 μM AgNO₃ and 1 mM FGE-27. **Bottom right:** Overlay of the spectra.

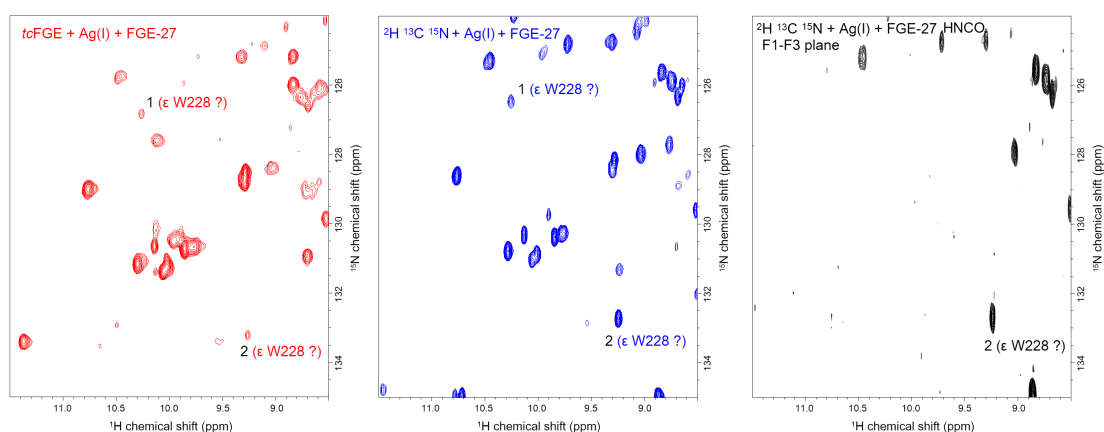


Figure S12 – NMR spectra of *tcFGE* in 20 mM phosphate buffer, 50 mM NaCl, 2 mM DTT, 2 mM glucose, 2 U GO, 200 U CAT, pH 8.1, 25°C. The downfield area of the amide region is presented. The ^1H - ^{15}N TROSY-HSQC experiment of triple-labelled ^2H ^{13}C ^{15}N *tcFGE* in complex with Ag(I) and FGE-27 was compared with the ^1H - ^{15}N TROSY-HSQC spectrum of ^{15}N labelled *tcFGE* containing Ag(I) and FGE-27. These spectra confirm that both cross peaks 1 and 2 are also visible in the ^1H - ^{15}N TROSY-HSQC experiment of triple-labelled *tcFGE* and none of them appeared just due to signal to noise issues. The subsequent comparison of ^1H - ^{15}N TROSY HSQC and F1-F3 plane from the HNCOC experiment for ^2H ^{13}C ^{15}N labelled *tcFGE* in the presence of Ag(I) and FGE-27 demonstrated, that that cross-peak **2** belongs to a ^1H - ^{15}N correlation either in the backbone or Asn/Gln side chain ³⁻⁴, whereas peak **1** is assigned as $^1\text{H}\epsilon 1$ - $^{15}\text{N}\epsilon 1$ correlation of W228. **Left:** ^1H - ^{15}N TROSY HSQC spectrum of ^{15}N labelled *tcFGE*_{W228F} 300 μM in the presence of 450 μM AgNO₃ and 1 mM of FGE-27. **Middle:** ^1H - ^{15}N TROSY HSQC spectrum of ^2H ^{13}C ^{15}N labelled *tcFGE* 500 μM in the presence of 750 μM AgNO₃ and 1.5 mM of FGE-27. **Right:** F1-F3 plane extracted from TROSY based HNCOC experiment for ^2H ^{13}C ^{15}N labelled *tcFGE* 500 μM in the presence of 750 μM AgNO₃ and 1.5 mM of FGE-27.

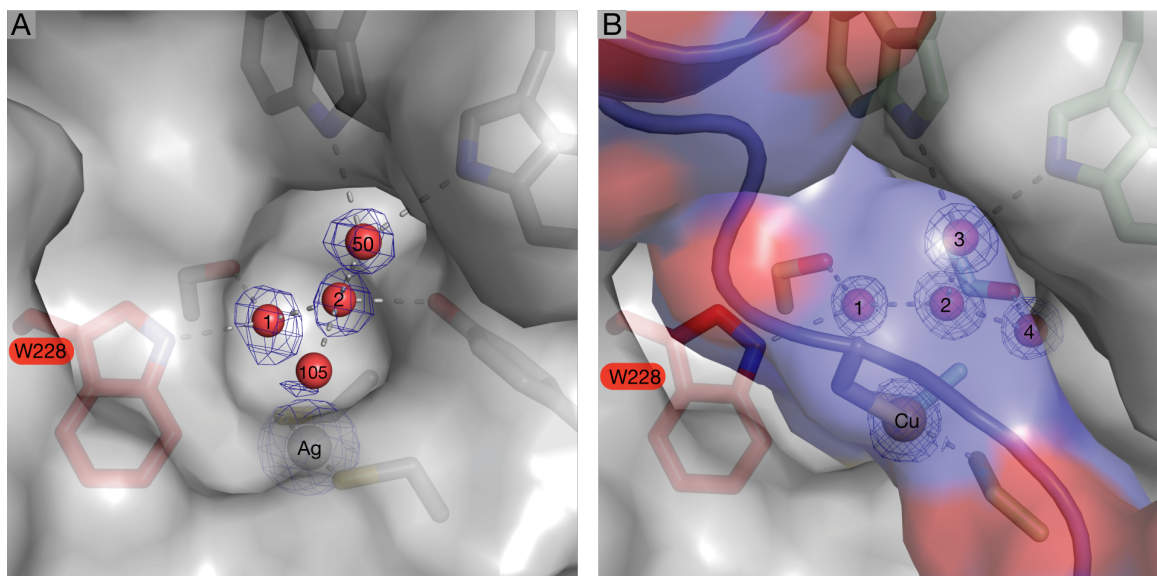


Figure S13 – The views into the active site of *tcFGE_Ag* (A) and *tcFGE_Cu_S* (B, substrate: violet) show that the water-exposed indol ring of Trp228 is buried upon substrate binding.

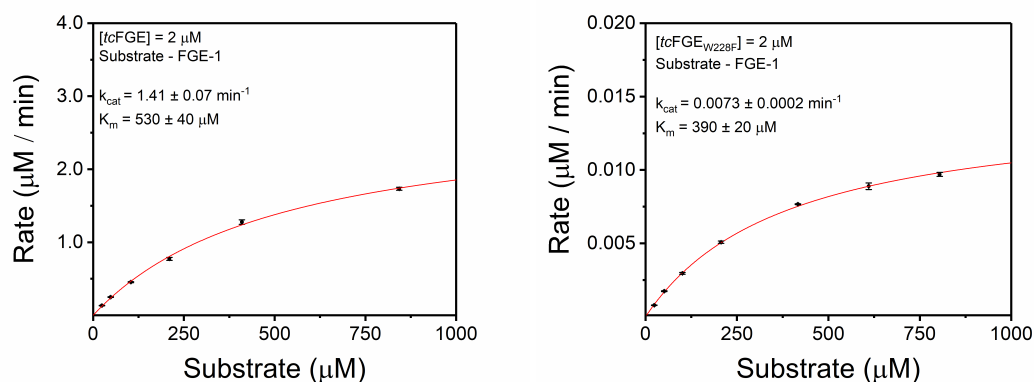


Figure S14 – Michaelis Menten kinetics of *tcFGE*- and *tcFGE_{W228F}*-catalyzed turnover of FGE-1 (Abz-SALCSPTRA-NH₂). The enzymes were assayed in reactions containing FGE-1 (25-1000 μM), 2 μM CuSO₄, 5 mM DTT, 50 mM NaCl and 50 mM Tris Buffer at pH 8.0, 25°C. Reaction aliquots were quenched by addition of 1 volume equivalent of 1.5 % TFA in 4 M urea. Reaction product were quantified by RP-HPLC (column Gemini-NX, 5 μm C18 250 x 4.6 mm, Phenomenex)⁵. Initial rates (v) from three independent measurements were fitted with the function $v = [S] \times k_{cat} / (K_M + [S])$ to obtain the Michaelis-Menten parameters k_{cat} and K_M .

SUPPLEMENTARY METHODS

Production of *tcFGE* variants for kinetic and ITC measurements. The nucleotide sequences encoding *tcFGE* variants were cloned into pET19b vector under the control *lac* promoter. The obtained plasmids were incorporated in *E.coli* BL21 DE3 pLyss to overexpress the proteins. The overnight cultures were inoculated into shaker flasks with 750 mL LB medium, supplemented with 100 mg/L Ampicillin and 37 mg/L Chloramphenicol, and were grown at 37°C, 180 rpm until OD₆₀₀ ~ 0.7. The protein expression was induced with 1 mM IPTG for 3 h. Then the cells were pelleted (9000 rpm, 4°C) and stored in the freezer (-20°C). The resuspended in lysis buffer (50 mM phosphate buffer pH 8.0, 300 mM NaCl, 10% Glycerol) cell pellet was lysed by sonication (Sonifer 450, Branson). The cleared lysate (25000 rpm, 4°C, 1 h) was treated with Ni(II)-NTA agarose (Quiagen) according to the standard affinity chromatography protocol. The purified proteins were dialyzed into 50 mM Tris-HCl buffer pH 8.0, 100 mM NaCl, flash frozen in liquid N₂ and stored at -80°C.

Production of ¹⁵N labelled *tcFGE* variants. The bacterial strains were cultivated in the fermenter INFORS HT in M9 medium, supplemented with ¹⁵NH₄Cl (99% ¹⁵N) as a source of nitrogen atoms. A single *E. coli* colony was inoculated to 5 mL M9 medium, supplemented with 1 mg/mL ¹⁵NH₄Cl, 3 mg/mL glucose, 100 mg/L ampicillin and 37 mg/L chloramphenicol, and incubated overnight at 37°C. The obtained preculture was transferred to 1 L M9 minimal medium, supplemented with 1 g ¹⁵NH₄Cl and 1 g glucose. The main culture was grown to OD₆₀₀ ~ 2.3 (~10 h, 37°C, 300 rpm, pH 7.2) and supplemented with 20 mL the feed solution (7.5 g Glc, 0.5 g ¹⁵NH₄Cl in 100 mL H₂O). Then the protein expression was induced with 1 mM IPTG at 20°C for 20 h. During the expression, the culture was fed with the rest of the feed solution. The consumption of glucose and ¹⁵NH₄Cl was monitored with QUANTOFIX® test stripes (MACHEREY-NAGEL). The harvested culture (OD₆₀₀ ~ 7.2-8.0) were centrifuged (9000 rpm, 20 min, 4°C) and stored in the freezer (-20°C). The resuspended in lysis buffer (50 mM phosphate buffer pH 8.0, 300 mM NaCl, 10% Glycerol) cell pellet was lysed with microfluidizer Emulsiflex HC3. The cleared lysate (25000 rpm, 4°C, 1 h) was treated with Ni(II)-NTA agarose (Quiagen) according to the standard affinity chromatography protocols. The obtained protein sample was treated with TEV-protease (0.05 w/w) during the dialysis against 100 mM phosphate buffer pH 8.0, 200 mM NaCl, 1 mM DTT at RT overnight. Then enzyme was further purified with size exclusion chromatography (HiLoad 26/600 Superdex 200 pg, 200 mM phosphate buffer pH 8.0, 500 mM NaCl). The obtained fractions concentrated with Amicon Ultra (15 mL centrifugal filter a membrane NMWL of 10 kDa), flash frozen in liquid N₂ and stored at -80°C.

Production of ²H ¹³C ¹⁵N labelled *tcFGE*. The bacterial strain was cultivated in the fermenter INFORS HT in the M9 medium, supplemented with D₂O (99.8% ²H), D-Glucose-¹³C₆²D₇ (97% ²D, 99% ¹³C) and ¹⁵NH₄Cl (99% ¹⁵N) as the sole of carbon and nitrogen atoms. A single colony *E.coli* BL21 DE3 pLyss bearing the plasmid pET19b_*tcFGE* was transferred to 10 mL M9 medium, supplemented with 1 mg/mL NH₄Cl, 3 mg/mL glucose, 100 mg/L ampicillin and 37 mg/L chloramphenicol, and incubated overnight at 37°C. The resulted culture was pelleted (8000 rpm, 5 min) and used to inoculate the shaker flask with 50 mL M9 D₂O medium, supplemented with 99.8% D₂O, 3 mg/mL D-Glucose-¹³C₆ ²D₇, 1 mg/mL ¹⁵NH₄Cl, 100 mg/L ampicillin and 37 mg/L chloramphenicol, which further was incubated overnight at 37°C, 200 rpm. The obtained cells (OD₆₀₀ ~ 0.8) were transferred in 11 M9 D₂O medium (99.8% D₂O, 1 g/l D-Glucose-¹³C₆²D₇, 0.5 g/l ¹⁵NH₄Cl, 100 mg/L ampicillin and 37 mg/L chloramphenicol). The main culture was grown to OD₆₀₀ ~ 2.3 (~21 h, 37°C, 300 rpm, pH 7.2) and supplemented with 15 mL the feed solution (7.0 g D-Glucose-¹³C₆²D₇, 0.5 g ¹⁵NH₄Cl, in 100 mL

D₂O). The protein expression was induced with 1 mM IPTG at 20°C for 10 h. During the expression, the culture was fed with the rest of the feed solution. The consumption of D-Glucose-¹³C₆D₇ and ¹⁵NH₄Cl was monitored with QUANTOFIX® test stripes (MACHEREY-NAGEL). The harvested culture (OD₆₀₀ ~ 8.6) were centrifuged (9000 rpm, 20 min, 4°C) and stored in the freezer (-20°C). The resuspended in lysis buffer (50 mM phosphate buffer pH 8.0, 300 mM NaCl, 10% Glycerol) cell pellet was lysed with microfluidizer Emulsiflex HC3. The cleared lysate (25000 rpm, 4°C, 1 h) was treated with Ni(II)-NTA agarose (Quiagen) according to the standard affinity chromatography protocols. The obtained protein sample was treated with TEV-protease (0.05 w/w) during the dialysis against 100 mM Tris-HCl buffer pH 8.0, 200 mM NaCl, 1 mM DTT at RT overnight. Then enzyme was further purified with size exclusion chromatography (HiLoad 26/600 Superdex 200 pg, 100 mM Tris-HCl buffer pH 8.0, 200 mM NaCl). To perform the backbone amide hydrogen exchange the protein sample was dialyzed against 40 mM Tris/HCl buffer pH 8.0, 150 mM NaCl in presence 1 mM EDTA at RT for 48 h with subsequent removal EDTA and incubation with 2 mM CaCl₂ (3h, RT). Afterwards the Tris buffer was exchanged to 200 mM phosphate buffer pH 8.0, 750 mM NaCl with Amicon Ultra (15 mL centrifugal filter a membrane NMWL of 10 kDa), concentrated, flash frozen in liquid N₂ and stored at -80°C.

Production of *tcFGE* for crystallization. The shaker flasks with 8*1 l ZYM-5052 ⁶ medium, supplemented with 100 mg/L Ampicillin and 37 mg/L Chloramphenicol, were inoculated with overnight culture of *E.coli* BL21 DE3 pLyss, bearing the plasmid pET19b_ *tcFGE*, and the cells were grown at 37°C, 180 rpm until OD₆₀₀ ~ 0.7. Afterwards the temperature was decreased to 20°C and the cultures were cultivated for 20 h. Then the cells were pelleted in the centrifuge (9000 rpm, 4°C) and stored in the freezer (-20°C). The resuspended in lysis buffer (50 mM phosphate buffer pH 8.0, 300 mM NaCl, 10% Glycerol) cell pellet was lysed with microfluidizer Emulsiflex HC3. The cleared lysate (25000 rpm, 4°C, 1 h) was treated with Ni(II)-NTA agarose (Quiagen) according to the standard affinity chromatography protocols. The obtained protein sample was treated with TEV-protease (0.05 w/w) during the dialysis against 100 mM phosphate buffer pH 8.0, 200 mM NaCl, 1 mM DTT at RT overnight. Then enzyme was further purified with size exclusion chromatography (HiLoad 26/600 Superdex 200 pg, 200 mM phosphate buffer pH 8.0, 500 mM NaCl). The obtained fractions concentrated with Amicon Ultra (15 mL centrifugal filter a membrane NMWL of 10 kDa).

The purity of the proteins was confirmed with SDS PAGE and HRMS (ESI). The labeling efficiency for the isotope labeled *tcFGE* variants was estimated on the base of HRMS results. The final protein concentration was estimated from the A₂₈₀ with usage of calculated from amino acid sequences molar absorption coefficients ⁷.

Crystallization and diffraction data collection. Sitting drop vapor diffusion method was applied for the crystallization of *tcFGE* under strict exclusion of dioxygen in an inert gas chamber containing 95/5 % mixture of N₂/H₂. Using an enzymatic oxygen scavenging system ⁸ dissolved oxygen was removed from protein and substrate sample. The protein sample *tcFGE* (21.5 mg/mL, 0.61 mM), CuSO₄ (0.83 mM), 2.7 mM DTT as well as FGE-27 (1.7 mM) was supplemented 2 mM glucose, 2 U glucose oxidase (0.01 mg/mL, *A. niger*, Sigma) and 200 U catalase (0.04 mg/mL, bovine liver, Sigma) and incubated in the glovebox (N₂ atmosphere, O₂ ≤ 0.1 ppm) overnight at RT. Prior to crystallization the protein substrate mixture was incubated for 90 min. On a 96-well 2-drop plate (SwissSci) 0.6 μL reservoir solution of crystallization screen was combined with 0.6 μL protein solution. Sealed plates were stored at 23°C for reservoir / protein equilibration and formation of three-dimensional crystals was observed after one week in a drop containing 10 % PEG 8000, Tris-HCl (0.1 M, pH 7.0) and MgCl₂ (0.2 M). Crystals were looped and flash frozen in liquid nitrogen approximately 3 weeks past

crystallization. Diffraction data were collected at X06DA (PXIII) and X06SA (PXI) beamline of Swiss Light Source (Paul Scherrer Institute, Villigen, Switzerland) equipped with a Pilatus 2 M (Dectris) or an Eiger-16M detector (Dectris), respectively. High resolution data sets were collected at an X-ray energy of 12398 eV ($\lambda = 1.000 \text{ \AA}$) at the temperature 100 K. Data were indexed and integrated with XDS⁹ and scaled and merged with Aimless¹⁰ from the CCP4 program suite¹¹⁻¹². *tcFGE* crystals grew into orthorhombic space group $P2_12_12_1$ with unit cell dimensions of $a = 58.4 \text{ \AA}$, $b = 71.9 \text{ \AA}$, $c = 76.7 \text{ \AA}$, $\alpha = 90^\circ$, $\beta = 90^\circ$, $\gamma = 90^\circ$. A single protein monomer was contained per unit cell.

Data processing and refinement. Due to the excellent diffraction quality the data were scaled to a high-resolution limit of 1.04 \AA . A model of the truncated N20FGE silver co-crystallized apo structure (PDB: 5NXL) was used as search model for molecular replacement using Phenix Phaser¹³. Iterative cycles of manual and automated model building into the readily interpretable electron density map and refinement using COOT¹⁴⁻¹⁵ and phenix.refine¹⁶⁻²⁰. Hydrogen atoms were generated and used for the refinement but not included into the final model. *tcFGE* polypeptide chain was modeled into continuous electron density for residue 1 – 303 including the so far unresolved residues from Val92 to Val102 with disorder for the C-terminal Leu303. Aside of the protein chain the asymmetric unit cell contained additional electron density that could have readily been modeled by adding the substrate polypeptide chain (NH₂-Abz-AlaThrThrProLeuCysGlyProSerArgAlaSerIleLeuSerGlyArg-CO₂H) resolved from N-terminal *o*-aminobenzoic acid to Gly17. Substrate peptide residues from Leu15 onwards are no more accommodated in a protein surface substrate binding cleft but are found in the solvent exposed area. More remarkable the ultimate C-terminal residues Gly11 & Arg12 are located at the interface of two crystallographic symmetry mates (-1,0,-1 & -1,-1,-1 molecules). The interaction of the substrate residues with further FGE molecules could be beneficial for crystallization. Data processing and refinement statistics are given in Table S3.

High-resolution mass spectrometry (HRMS). For HRMS measurements protein samples were diluted to concentration 0.2 mg/mL in 1 mM DTT, 1% FA and injected into a reversed-phase C4 column (Jupiter C4 300A; 50 x 2 mm, 5 Microns, Phenomex; Shimadzu Nexera-X2) coupled to a HRMS instrument (maXis II ESI-TOF, Bruker). The obtained MS data were deconvoluted with Bruker Compass DataAnalysis 4.4 SR1 software.

Solid phase peptide synthesis (SPPS). The substrate peptide 17-residue peptide (FGE-27, sequence: Abz-ATTPLCGPSRASILSGR-NH₂) and 9-residue peptide (FGE-1, Abz-SALCSPTRA-NH₂) were synthesized with Syro I peptide synthesizer (Biotage) using standard protocols for Fmoc SPPS.²¹ The crude product was purified to homogeneity with C18 reverse phase HPLC and analyzed with ESI-MS (Agilent 6130 Quadrupole LC/MS, C18 column ZORBAX Eclipse Plus RRHD, 1.8 μm , 2.1 x 50 mm) (Figure S17).

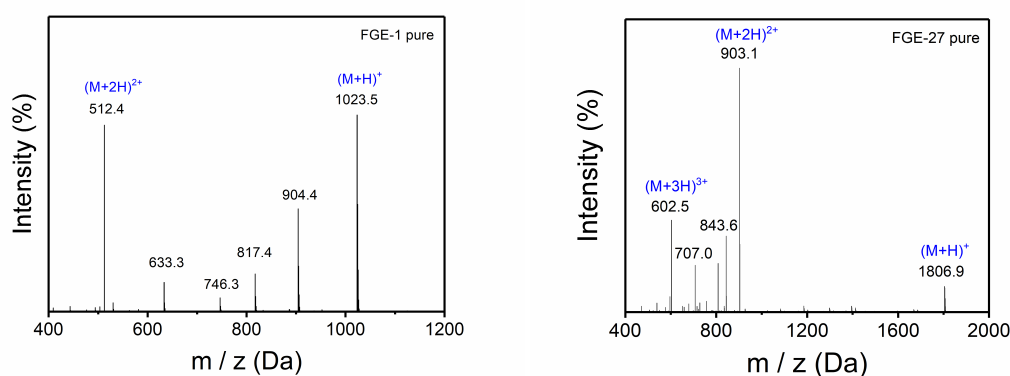


Figure S15 – The mass spectra of substrates after purification with C18 reverse phase HPLC. The analyses were performed with UPLC-MS Agilent 1290 Infinity system equipped with an Agilent 6130 Quadrupole LC/MS using a C18 column (ZORBAX Eclipse Plus RRHD, 1.8 μ m, 2.1 x 50 mm) **Left:** The mass spectra purified FGE-1 showed protonated ($m/z = 1023.5$ Da) and double protonated ($m/z = 512.4$ Da) peaks. **Right:** The mass spectra purified FGE-27 showed protonated ($m/z = 1806.9$ Da), double protonated ($m/z = 903.1$ Da) and triple protonated ($m/z = 602.5$ Da) peaks.

NMR experiments. NMR spectra of apo forms of *tcFGE* variants as well as metalated one were recorded in oxygen-contained buffers. To prevent substrate oxidation during NMR experiments dissolved oxygen was removed with an enzymatic oxygen scavenging system⁸. The protein as well as FGE-27 were separately diluted in 20 mM phosphate buffer (pH 8.2), supplemented with 50 mM NaCl, 2 mM DTT, 2 mM glucose, 2 U glucose oxidase (*A. niger*, Sigma) and 200 U catalase (bovine liver, Sigma) and 5% D₂O. After addition of the GO and CAT samples were incubated in the glovebox (N₂ atmosphere, O₂ \leq 0.1 ppm) overnight at RT. Just before every measurement, the enzyme and the substrate solutions were combined and 350 μ L of final solution were transferred to a SHIGEMI micro cell tube in the glovebox. TROSY based experiments were used to record ¹H-¹⁵N HSQC²² and triple resonance 3D-HNCO²³ spectra. NMR spectra were recorded on Bruker Avance III NMR spectrometers operating at 600.13 MHz, either equipped with a 5 mm BBFO probe head with z-axis pulsed field gradients or a 5 mm ¹H/¹³C/¹⁵N/¹⁹F QCI-F cryogenic probe head with z-axis pulsed field gradients.

SUPPLEMENTARY INFORMATION REFERENCES

1. Meury, M.; Knop, M.; Seebeck, F. P., Structural Basis for Copper-Oxygen Mediated C-H Bond Activation by the Formylglycine-Generating Enzyme. *Angew Chem Int Ed Engl* **2017**, *56* (28), 8115-8119.
2. Mollova, E. T.; Metzler, D. E.; Kintanar, A.; Kagamiyama, H.; Hayashi, H.; Hirotsu, K.; Miyahara, I., Use of ¹H-¹⁵N heteronuclear multiple-quantum coherence NMR spectroscopy to study the active site of aspartate aminotransferase. *Biochemistry* **1997**, *36* (3), 615-25.
3. John Cavanagh, W. J. F., Arthur G. Palmer III, Mark Rance and Nicholas J. Skelton *Protein NMR Spectroscopy - Principles and Practice*. ACADEMIC PRESS: 2007.
4. Muhandiram, D. R.; Kay, L. E., Gradient-Enhanced Triple-Resonance Three-Dimensional NMR Experiments with Improved Sensitivity. *Journal of Magnetic Resonance, Series B* **1994**, *103* (3), 203-216.
5. Knop, M.; Engi, P.; Lemnar, R.; Seebeck, F. P., In Vitro Reconstitution of Formylglycine-Generating Enzymes Requires Copper(I). *Chembiochem* **2015**, *16* (15), 2147-50.
6. Studier, F. W., Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification* **2005**, *41* (1), 207-234.
7. Gasteiger E., H. C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A., Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook*, Walker, J. M., Ed. Humana Press: 2005; pp 571-607.
8. Plumere, N.; Henig, J.; Campbell, W. H., Enzyme-catalyzed O₂ removal system for electrochemical analysis under ambient air: application in an amperometric nitrate biosensor. *Anal Chem* **2012**, *84* (5), 2141-6.
9. Kabsch, W., Xds. *Acta Crystallogr D Biol Crystallogr* **2010**, *66* (Pt 2), 125-32.
10. Evans, P. R.; Murshudov, G. N., How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr* **2013**, *69* (Pt 7), 1204-14.
11. Evans, P., Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* **2006**, *62* (Pt 1), 72-82.
12. Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G.; McCoy, A.; McNicholas, S. J.; Murshudov, G. N.; Pannu, N. S.; Potterton, E. A.; Powell, H. R.; Read, R. J.; Vagin, A.; Wilson, K. S., Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **2011**, *67* (Pt 4), 235-42.
13. McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J., Phaser crystallographic software. *J Appl Crystallogr* **2007**, *40* (Pt 4), 658-674.
14. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K., Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **2010**, *66* (Pt 4), 486-501.
15. Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **2004**, *60* (Pt 12 Pt 1), 2126-32.
16. Afonine, P. V.; Grosse-Kunstleve, R. W.; Echols, N.; Headd, J. J.; Moriarty, N. W.; Mustyakimov, M.; Terwilliger, T. C.; Urzhumtsev, A.; Zwart, P. H.; Adams, P. D., Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* **2012**, *68* (Pt 4), 352-67.
17. Terwilliger, T. C., Using prime-and-switch phasing to reduce model bias in molecular replacement. *Acta Crystallogr D Biol Crystallogr* **2004**, *60* (Pt 12 Pt 1), 2144-9.
18. Chen, V. B.; Arendall, W. B., 3rd; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C., MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **2010**, *66* (Pt 1), 12-21.
19. Headd, J. J.; Echols, N.; Afonine, P. V.; Grosse-Kunstleve, R. W.; Chen, V. B.; Moriarty, N. W.; Richardson, D. C.; Richardson, J. S.; Adams, P. D., Use of knowledge-based restraints in phenix.refine to improve macromolecular refinement at low resolution. *Acta Crystallogr D Biol Crystallogr* **2012**, *68* (Pt 4), 381-90.
20. Terwilliger, T. C.; Grosse-Kunstleve, R. W.; Afonine, P. V.; Moriarty, N. W.; Adams, P. D.; Read, R. J.; Zwart, P. H.; Hung, L. W., Iterative-build OMIT maps: map improvement by iterative model building and refinement without model bias. *Acta Crystallogr D Biol Crystallogr* **2008**, *64* (Pt 5), 515-24.
21. W. CHAN, P. W., *Fmoc solid phase peptide synthesis A Practical Approach*. Oxford University Press Inc.: New York, 2000.
22. Pervushin, K.; Riek, R.; Wider, G.; Wüthrich, K., Attenuated T₂ relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proceedings of the National Academy of Sciences of the United States of America* **1997**, *94* (23), 12366-12371.
23. Salzman, M.; Pervushin, K.; Wider, G.; Senn, H.; Wüthrich, K., TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. *Proceedings of the National Academy of Sciences of the United States of America* **1998**, *95* (23), 13585-13590.

6.2 The molecular basis for multiple sulfatase deficiency caused by mutations in the outer calcium binding site of formylglycine-generating enzyme

This project was performed in collaboration with Dzmitry Miarzlou and Prof. Dr. Florian P. Seebeck from the University of Basel.

6.2 The molecular basis for multiple sulfatase deficiency caused by mutations in the outer calcium binding site of formylglycine-generating enzyme

6.2.1 Abstract

Formylglycine-generating enzyme stands out as a unique copper-dependent protein that controls the activation of type I sulfatases. The absence or malfunction of formylglycine-generating enzyme leads to the severe, inherited disease termed multiple sulfatase deficiency. Herein, the unusual structure of the apo- and holoenzyme as well as two disease-related mutants that were identified in patients, N188S and N188I, were characterized by nuclear magnetic resonance spectroscopy, mass spectrometry and biophysical techniques. Thereby, it is shown that the molecular basis for multiple sulfatase deficiency constitutes a 16- (N188S) respectively 32-fold decrease (N188I) decrease of copper affinity and the loss of stability of the protein fold. Furthermore, the analysis of NMR relaxation data reveals a stabilizing network between the calcium and copper binding sites in the wild type fold.

6.2.2 Introduction and aim of the study

Stability of proteins in solution is crucial to the function and survival of enzymes.¹ Proteins with a stable and rigid fold in solution typically show a substantial percentage of residues in secondary structural elements, as e.g. alpha-helices or beta-sheets. In order to fulfil the important catalytic and regulatory functions of enzymes, a strictly defined structure is the essential feature for stabilization of the transition state and hence catalytic activity.²⁻³ Formylglycine-generating enzyme (FGE) displays several interesting structural properties. While common catalytically active copper proteins display a set of N-containing ligands⁴⁻¹⁰ and copper trafficking proteins contain a bislinear cysteine coordination motif,¹¹⁻¹² FGE is to date the only observed highly catalytically active copper protein bearing two cysteines coordinating the Cu(I) cofactor in bislinear geometry.¹³ Furthermore, it is visible from the X-ray structure that FGE contains only few classical secondary structure elements, i.e. alpha-helices and beta-sheets. In contrast to the intuitive assumption that this should lead to a decreased stability of the enzyme, the protein fold of FGE shows a very high temperature stability of 80 °C. FGE catalyzes the oxygen-dependent conversion of specific cysteine residues of a protein substrate to formylglycine using Cu(I) as cofactor.¹⁴⁻¹⁶ Importantly, a lack of FGE present in humans can lead to the severe disease multiple sulfatase deficiency (MSD), caused by the absence of activation of different sulfatases.¹⁷⁻²⁰ Recently, a structural characterization of a FGE-substrate complex as well as the preactivated oxygen complex was provided.^{13, 21} An intriguing question that remains, is how the fold of FGE is affected by disease-related mutations as N188S and

N188I that were identified in MSD patients.²²⁻²³ Mutations in the active site of FGE clearly perturb the activity and hence lead to the inability of FGE to activate sulfatases. However, the N188S and N188I mutations are located in the outer Ca(II) binding site and thus the rationalization of their impact on the function of FGE is not straightforward.

In order to investigate and study the exceptional protein fold of wild type FGE (FGEwt) and the two disease-related mutants in solution, we analyzed and compared the Ramachandran plot²⁴ to commonly structured, folded, globular proteins²⁵⁻²⁷ and analyzed amide hydrogen/deuterium exchange observed by NMR spectroscopy. Furthermore, ¹⁵N T₁ and T₂ relaxation times as well as heteronuclear nuclear Overhauser effects (heteronuclear NOEs) were analyzed in order to get more insights into the proteins structure and its crucial stabilization factors and networks in solution. Together with chemical shift perturbations (CSPs) in ¹H-¹⁵N heteronuclear single quantum coherence (HSQC) experiments and the characterization of FGE by mass spectrometry (MS), melting points and kinetic experiments, new insights into the molecular basis of MSD were obtained.

6.2.3 Results and discussion

The structure of FGE incorporates an amount of 23.3% charged amino acids (average in Uniprot database: 25.8%) and it exhibits a high number of 10 salt bridges with an average bridged distance of 2.9 Å within the protein. 6 out of the 10 salt bridges are conserved among proteins from different species sharing the FGE fold (Figure S1, S2, Table S1). Despite the rather low number of residues in secondary structural elements within the core and around the active site, FGE adopts a compact, globular fold, stabilized by H-bonding networks, salt bridges, hydrophobic interactions and three metal ions (Figure 1).

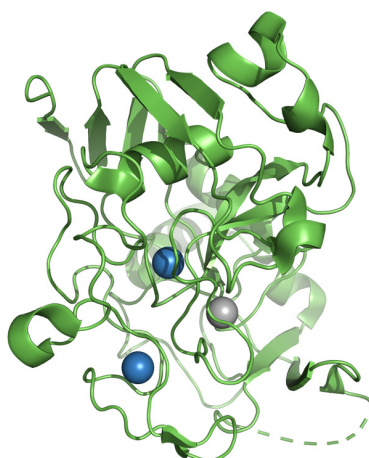


Figure 1: Crystal structure of the FGE protein in complex with a Ag(I) ion (silver) in its native state incorporating two Ca(II) ions (blue) (PDB 5NXL¹⁶).

FGE was readily crystallized as holoenzyme containing Ag(I) or Cd(II) as well as Cu(I)-containing protein in complex with a peptide substrate.^{13, 16}

In order to get an insight into the dihedral angles present in FGE and thus perform a conformational analysis, we analyzed the Ramachandran plots for three different proteins together with formylglycine generating enzyme from *T. curvata* in complex with Ag(I) (Figure 2).^{16, 25-27} From the Ramachandran plot depicted in Figure 2, it is obvious that the dihedral angles present in the FGE protein are comparable to globular proteins with an amount of secondary structural elements above 60% as for example cobalamin-complexed bovine transcobalamin or deoxy hemoglobin.

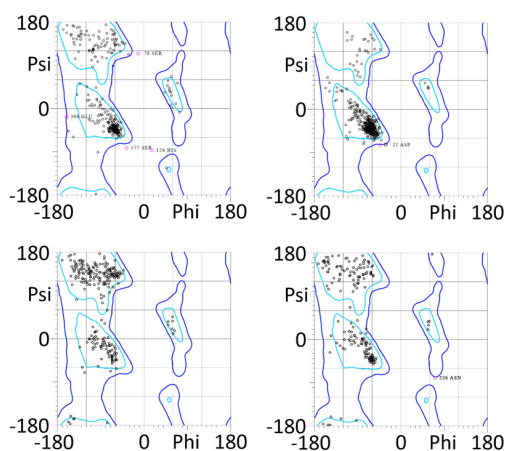


Figure 2: Ramachandran plots of cobalamin-complexed bovine transcobalamin (PDB 2BBC²⁵, upper left), deoxy hemoglobin (PDB 1HGA²⁶, upper right), porcine pancreatic trypsin (PDB 1AVW²⁷, lower left) and FGE from *T. curvata* in complex with Ag(I) (PDB 5NXL¹⁶, lower right).

However, since FGE does not exhibit an extensive amount of secondary structural elements (33%), we wondered how the found range of dihedral angles in FGE are reflected in a sequence analysis by STRIDE.²⁸ Therefore, we investigated the adopted dihedral angles of all residues and the associated propensity for formation of secondary structural elements. This analysis yields that, although FGE possesses dihedral angles comparable to proteins with a high content of extended secondary structure elements, the alternating pattern of residues inhibits the protein backbone from adopting extended secondary structural elements. This property manifests therefore in alternating turn and beta-strand/beta-bridge motifs (Figure S3). In order to investigate the stability of FGE in the apo- and holo-form as well as the two disease-related mutants N188S and N188I, we observed the amide H/D exchange in solution for wild type FGE with and without Cu(I) as well as the N188S and N188I variants in complex with Cu(I) by NMR. The global amide exchange behavior observed in our NMR experiments reflects the stabilities of the different FGE variants and leads to the conclusion, that the investigated disease-related mutations cause a destabilization of the protein fold. Most importantly, the

N188S variant partially destabilizes the FGE fold and N188I leads to a severe destabilization of the entire fold, which manifests in a rapid exchange of nearly all tracked signals in ^1H - ^{15}N HSQC experiments (Figure 3). The residues that remain unexchanged in FGE N188I show strong H-bonding interactions.

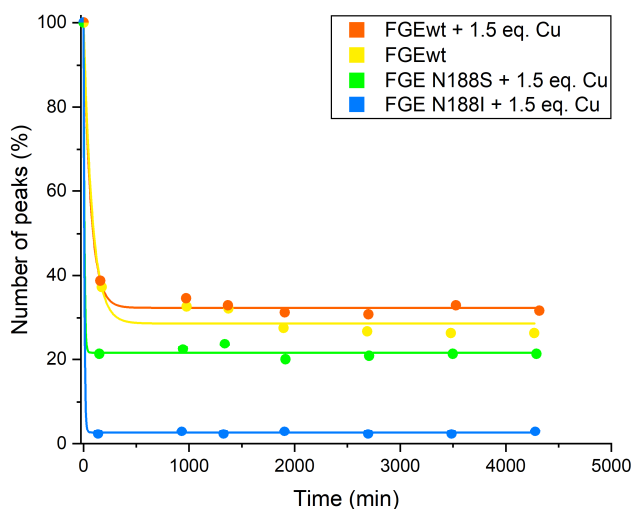


Figure 3: Percentage of non-exchanged residues for wild type FGE with and without Cu(I) as well as for the two disease-related mutants N188S and N188I in complex with Cu(I). The stability drops slightly for wild type FGE without Cu(I) when compared to its variant with Cu(I). The two disease-related mutations cause a significant destabilization of the protein fold.

For the wild type enzyme with Cu(I), a significant portion of ~32% of the residues remain visible in ^1H - ^{15}N HSQC experiments even after 3 d (Figure 3, Table S6), thus confirming the strong stabilization of the FGE fold that manifests in a melting point of around 80 °C. Interestingly, the enzyme without Cu(I) shows only a weakly destabilized fold (Figure 3, Table S7). However, for the disease-related FGE N188S and N188I variants a clear destabilization of the fold is observed and 21% (N188S) and 3% (N188I) of the residues remain after 3 d (Figure 3, Table S8 and S9).

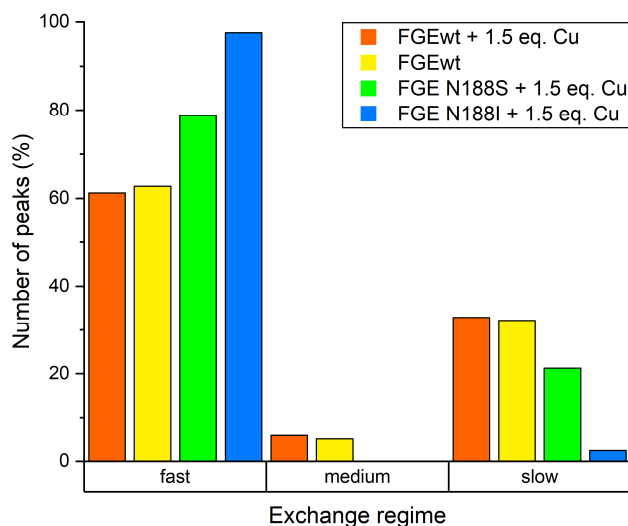


Figure 4: Percentage of fast, medium and slowly exchanging peaks in the different investigated variants (fast exchange: < 3 h, medium exchange: < 1 d, slow exchange: remains > 1 d). For the destabilized variants FGE N188S and N188I, a clear simultaneous increase in fast exchanging peaks and decrease in slow exchanging peaks is visible when compared to FGEwt + 1.5 eq. Cu(I).

In order to investigate the exchange behavior of the different FGE variants in more detail, we separated the vanishing of peaks in three groups (fast exchange: < 3 h, medium exchange: < 1 d, slow exchange: remains > 1 d). The analysis yields a significant increase for the percentage of fast exchanging peaks and a strong decrease of slowly exchanging peaks for the destabilized variants N188S and N188I (Figure 4, Table S10) in line with the global amide exchange plot (Figure 3, Table S6-9). The disease-related mutant N188S shows only 65% of the amount of slowly exchanging backbone amides when compared with the wild type FGE incorporating Cu(I). Most strikingly, the amount of slowly exchanging peaks decreases very sharply for the N188I variant, that shows only 7% of the amount observed for wild type FGE with Cu(I). Both the N188S and the N188I variant display a strong rise in the amount of fast exchanging amides in their fold when compared to wild type FGE with Cu(I) (N188S: +29%, N188I: +59%). The melting points measured for the different FGE variants confirm the stabilities obtained in the amide exchange experiments (Table 1).

Table 1: The melting points of apo- and holo-FGE variants determined with differential scanning fluorimetry (DSF).

Variant	Melting temperature (°C)
FGEwt with 1.0 eq. Cu(I)	82.14 ± 0.08
FGEwt	67.34 ± 0.05
FGE N188S with 1.0 eq. Cu(I)	53.40 ± 0.10
FGE N188I with 1.0 eq. Cu(I)	50.74 ± 0.02

Interestingly, the MS analysis leads to the finding that the mutant N188S contains only one Ca(II) ion and N188I lost both of the Ca(II) ions present in the wild type fold of FGE (Figure S4, Table S2). In wild type FGE, the sidechain amide group of asparagine 188 serves as ligand for complexation of the Ca(II) ion. Due to the apolar nature and the absence of the amide group in the isoleucine sidechain, this interaction is expected to be lost in the N188I variant. For the N188S variant with its polar, but one bond shorter sidechain when compared to asparagine, the coordination of the Ca(II) ion by the sidechain hydroxy group of the serine should be possible upon minimal reorganization of the outer Ca(II) binding site. Since MS analysis indicates the loss of one of the two Ca(II) ions for the N188S variant, we resided to NMR experiments in order to clarify, if the inner or outer Ca(II) ion was lost in FGE N188S. An NMR backbone assignment was performed at 310 K for both the uniformly ^2H , ^{13}C , ^{15}N -labeled variants of the Cu(I) enzyme comprising two Ca(II) ions and the Ag(I) variant, that lacks the inner Ca(II) ion. The assignments at 310 K were transferred by a temperature series to 298 K. For the disease-related mutant FGE N188S, weak CSPs for the measured residues (Table S3, Figure 5) in the outer Ca(II) binding site revealed that the outer Ca(II) ion is present. This is in agreement with the obtained MS data that indicates the presence of one Ca(II) ion.

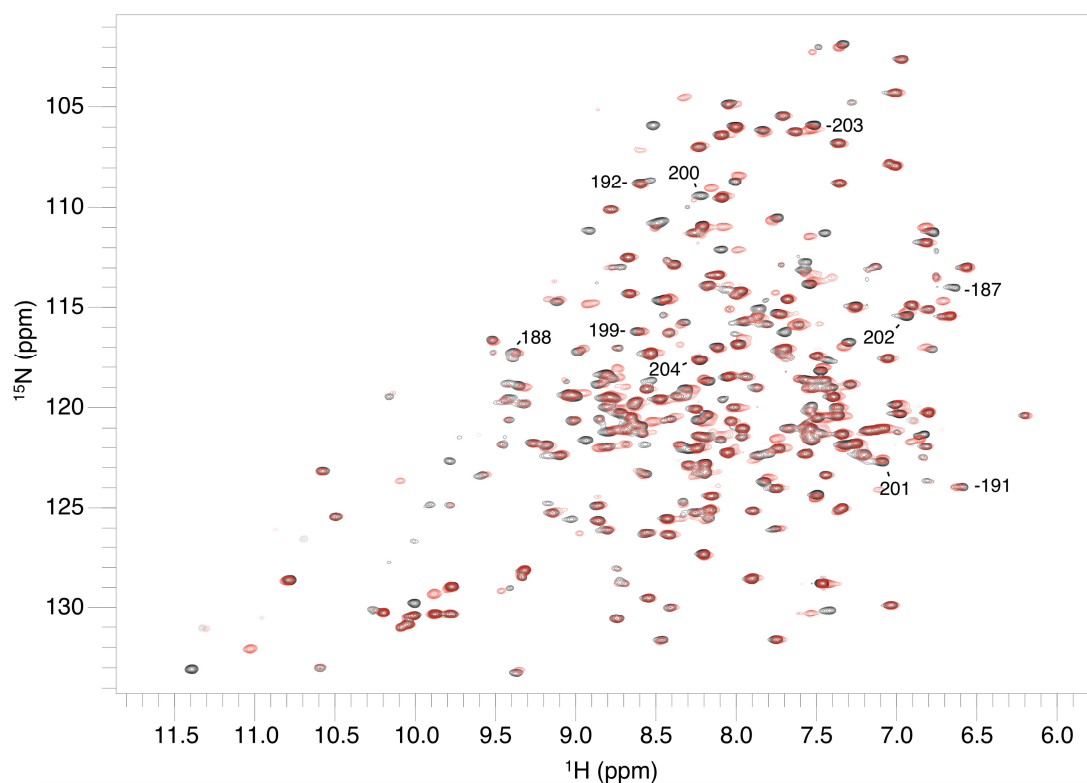


Figure 5: Overlay of ^1H - ^{15}N TROSY-HSQC spectra of ^{15}N FGEwt with Cu(I) (black) and ^{15}N FGE N188S with Cu(I) (red). Small CSPs are found for the residues in the outer Ca(II) binding site of FGE N188S.

Due to the weak CSPs for the outer Ca(II) binding site of FGE N188S in ^1H - ^{15}N HSQC experiments that confirm the presence of the outer Ca(II) ion, and the MS experiments that yield the presence of only one Ca(II) in the FGE N188S variant, the Ca(II) ion in the inner Ca(II) binding site has to be absent in FGE N188S.

In order to obtain further and direct evidence for the absence of the Ca(II) ion in the inner Ca(II) site of the disease-related variant N188S, ^1H - ^{15}N HSQC spectra of uniformly ^{15}N -labeled wild type FGE with Cu(I) and FGE N188S with Cu(I) were overlaid (Table S4, Figure S6).

The amide deuterium atoms in the inner Ca(II) binding site of the uniformly $^2\text{H}^{15}\text{N}^{13}\text{C}$ -labeled FGE with Cu(I) that contains two Ca(II) ions are not back-exchanged to hydrogen atoms during the back-exchange after the expression in D_2O . This result reflects the high gain in stability for the protein scaffold upon inclusion of the inner Ca(II) ion, but limits the assignment of the amide protons in the inner Ca(II) binding site.

However, the signals that were assigned to the inner calcium binding site in FGE N188S exhibit either a significant CSP or are not traceable (Table S4, Figure S6) in the ^{15}N -labeled wild type FGE sample with Cu(I). The large CSPs and/or the inability to trace the signals of the inner Ca(II) binding site in ^1H - ^{15}N HSQC experiments confirm the absence of the inner Ca(II) ion in the disease-related mutant N188S.

To summarize, the data obtained by NMR and MS shows that the disease-related variant N188S contains only a Ca(II) ion in the outer Ca(II) binding site, whereas the Ca(II) ion in the inner Ca(II) binding site is absent.

Concerning the FGE N188I variant, it was found that both the signals in the outer and inner Ca(II) site are shifted severely and could not be assigned with our data set (Figure 6). The closest candidates for the outer Ca(II) binding site of FGE N188I with Cu(I) in comparison with wild type FGE are given in Table S5.

The loss of both Ca(II) ions detected by CSPs in ^1H - ^{15}N HSQC experiments is in line with the MS data that indicates the absence of both Ca(II) ions within the fold of FGE N188I.

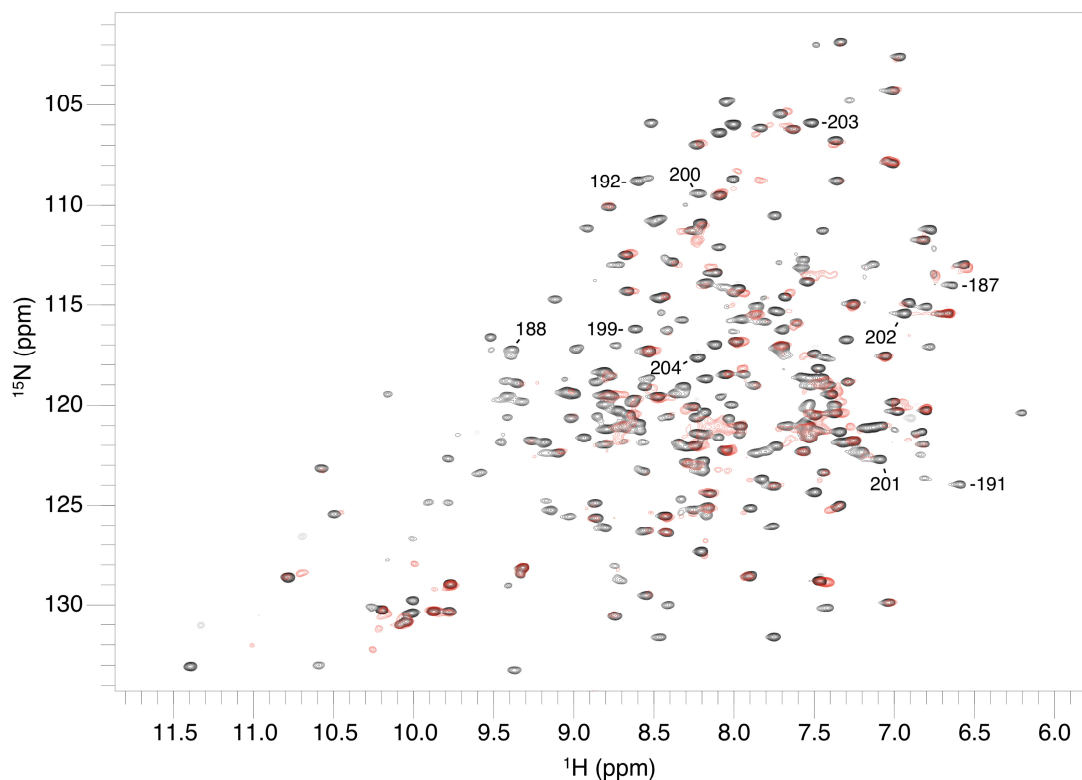


Figure 6: From the comparison of ^{15}N FGEwt with Cu(I) (black) and ^{15}N FGE N188I with Cu(I) (red) it is visible, that all signals corresponding to the outer Ca(II) binding site exhibit either large CSPs or are not traceable in the ^{15}N FGE N188I with Cu(I) sample.

To conclude, the absence of one Ca(II) ion in FGE N188S respectively two Ca(II) ions in FGE N188I clearly explains the amount of destabilization observed in the amide H/D exchange and melting point experiments for the two disease-related mutants.

Upon elucidation of the structural factors leading to a defective FGE, i.e. the loss of one or two Ca(II) ions of its scaffold, the question emerged, if these changes directly affect the catalytic properties of the enzyme or what other factor causes the absence of sulfatase activation and thus the final occurrence of MSD. Importantly, we found that all FGE variants are able to catalyze the conversion of a cysteine residue to a formylglycine with catalytic rates *in vitro* (Figure 7, Table S11) that should be sufficient for FGE function *in vivo*.²⁹ Therefore, the loss of one or two Ca(II) ions and the associated destabilization does only moderately perturb the ability of the FGE variants in complex with Cu(I) to catalyze the conversion of a cysteine to a formylglycine residue and cannot explain the severe MSD phenotypes.

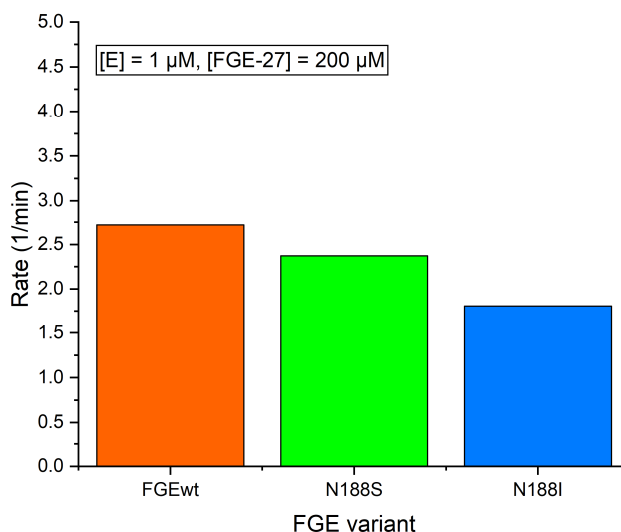


Figure 7: Kinetic measurements for the different investigated FGE variants confirming a change in rate for the disease-related mutants that is not sufficient to explain severe MSD phenotypes²⁹ (FGE-27: peptide substrate, 17 amino acids, Table S12). The rate was normalized to the enzyme concentration.

Importantly, measurement of the Cu(I) affinity for both disease-related mutants led to the finding that the ability to bind Cu(I) is significantly impaired for both the N188S and N188I mutant. While the N188S variant shows a 16-fold drop in Cu(I) affinity, the N188I mutant leads to a 32-fold decrease (Table 2).

Table 2: Cu(I) affinity of holo-FGE variants measured with a competition assay established for FGE.¹⁵

Variant	K _D
FGEwt	2.5 • 10 ⁻¹⁸
FGE N188S	4.0 • 10 ⁻¹⁷
FGE N188I	7.9 • 10 ⁻¹⁷

The severely lower Cu(I) affinity could explain the significant perturbation of *in vivo* catalysis observed for the two disease-related mutants and provides together with the decreased stability a molecular basis for FGE malfunction that results in MSD.

Subsequently, the backbone assignment was used in order to get a residue specific analysis of ¹⁵N T₁ and T₂ relaxation as well as heteronuclear NOEs on the uniformly ²H, ¹³C, ¹⁵N-labeled samples. Both the variant with one and the variant with two Ca(II) ions show a relative rigidification of the Cu(I)/Ag(I) and the outer Ca(II) site. Furthermore, it was found that the empty inner Ca(II) binding site of the FGE variant with Ag(I), that only incorporates the outer Ca(II) ion, appears to be only moderately rigidified. For the FGE variant with both the inner and outer Ca(II) ion present, the inner Ca(II) binding site was not assigned due to a lack of back-exchange after expression. This renders it impossible to measure the flexibility of the inner

Ca(II) binding site with Ca(II) ion present, however, the lack of back-exchange is associated with a severe, local stability gain when comparing the species incorporating both Ca(II) ions with the species with only the outer Ca(II) ion. Interestingly, the ^{15}N T_1 , T_2 and heteronuclear NOE data sets do not show significant differences in global flexibility for the species with 1 or 2 incorporated Ca(II) ions. The overall T_2 relaxation time for the 35 kDa FGE (1 Ca: $T_2=37$ ms; 2 Ca: $T_2=41$ ms) agrees with the experimental and calculated values of Venters et al. for the 29 kDa protein hCA II in ^2H , ^{13}C , ^{15}N -isotope labeled form ($T_{2,\text{calc}} = 49$ ms, $T_{2,\text{exp}} = 52$ ms).

6.2.4 Conclusions and Outlook

To conclude, the unusual fold of FGE and the molecular basis of MSD were investigated by NMR spectroscopy, MS and biophysical techniques. Importantly, a severe destabilization of the disease-related N188S and N188I was detected in amide H/D exchange experiments and confirmed by the melting points of the native and the disease-related species. MS analysis showed that the destabilization originates in a loss of one (FGE N188S) or two (FGE N188I) Ca(II) ions within the scaffold. Most importantly, NMR spectroscopy proved that the inner Ca(II) of N188S was lost and thus, a destabilization of the tertiary structure of the protein is inevitable. The biophysical studies led to the conclusion that the kinetics of the catalyzed reaction itself are not strongly hampered *in vitro* for the disease-related variants N188S and N188I. However, the associated 16- (N188S) respectively 32-fold decrease (N188I) in the affinity of the scaffold towards Cu(I) and the overall destabilization of the protein fold are most likely connected to the severely perturbed function of these two FGE variants *in vivo* and the occurrence of MSD. Besides the thorough investigation of the disease-related mutations, the FGE fold of the wild type species with one and two Ca(II) ions was characterized by NMR relaxation data, which highlighted the rigidification of the FGE core network around the three metal binding sites. Furthermore, based on the alternating patterns of dihedral angles, the unusual structure of FGE can be rationalized. In future, the characterization of further malfunctioning FGE variants by ^1H - ^{15}N HSQC experiments can be performed based on the NMR assignment and thus could lead ultimately to the discovery of a new, potent medication for MSD.

6.2.5 Supporting information



Figure S1: Sequence alignment of FGE from different organisms (D. Miarzlou).



Figure S2: Sequence alignment of proteins sharing the FGE fold (D. Miarzlou).

- tcFGE** formylglycine generating enzyme, *Thermomonospora curvata* (PDB 5NXL, 6S07)^{13, 16}
- hsFGE** formylglycine generating enzyme SUMF1, *Homo sapiens* (PDB 1Y1E)³⁰
- paPvdo** chromophore maturation protein, *Pseudomonas aeruginosa* (PDB 5HHA)³¹
- rsNirV** nitrate reductase, *Rhodobacter sphaeroides*³²
- chthEgtB** hercynylcysteine S-oxide synthase, *Chloracidobacterium thermophilum* (PDB 6QKI)³³
- mthEgtB** γ-glutamyl hercynylcysteine S-oxide synthase, *Mycobacterium thermoresistibile* (PDB 4X8D)³⁴
- sCarF** carbapenem intrinsic resistance protein, *Serratia sp. ATCC 39006* (PDB 5AOH)³⁵
- bsXylFR** transcriptional regulator of xylanase, *Bacillus stearothermophilus O31397*
- bsYfmG** uncharacterized protein, *Bacillus subtilis subsp. subtilis str. 168*

Table S1: Conserved salt bridges in FGE.

Positively charged residue	Negatively charged residue	Positioning in scaffold	Mode of conservation	Conserved in	Distance (Å)
Arg130	Glu121	surface	conserved	FGE	2.9
Lys152	Asp57	surface	conserved	FGE	3.2
Arg153	Glu159	surface	conserved	FGE fold	2.8
Arg153	Asp301	surface	conserved	FGE	3.0
Arg165	Glu161	surface </td <td>conserved</td> <td>FGE fold</td> <td>3.1</td>	conserved	FGE fold	3.1
Lys263	Glu157	buried	conserved	FGE fold	2.9
Arg276	Asp41	surface	conserved	FGE	2.8
Arg278	Glu43	surface	conserved	FGE fold	2.9
Arg282	Glu157	buried	conserved	FGE fold	2.9
Arg297	Glu229	buried	conserved	FGE fold	2.8

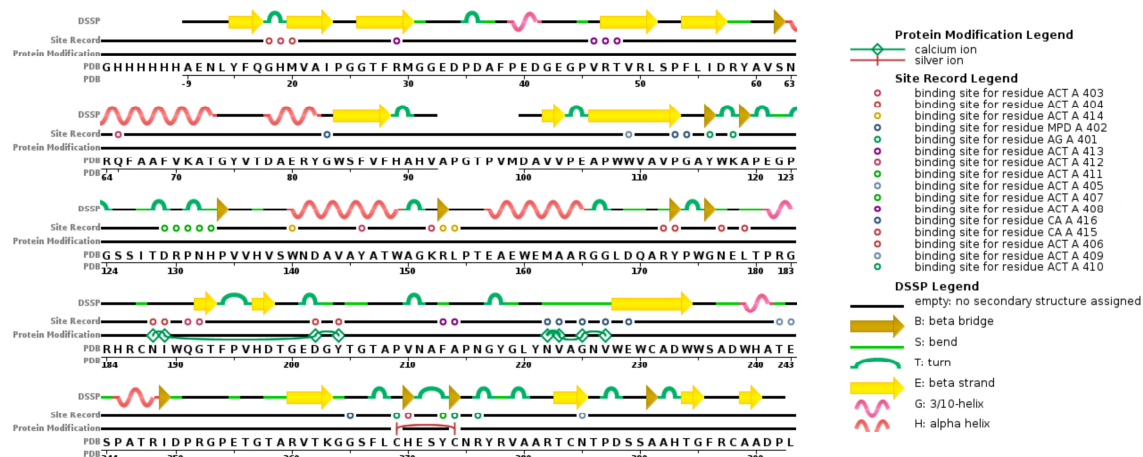


Figure S3: Sequence analysis of FGE incorporating Ag(I) by STRIDE²⁸ (PDB 5NXL¹⁶).

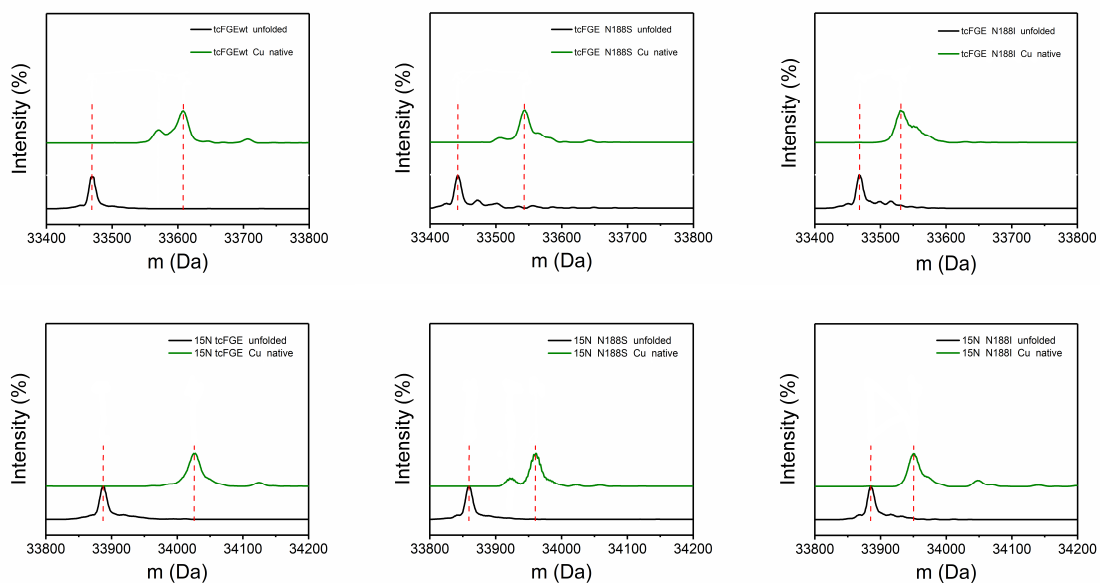


Figure S4: The Ca(II) metalation pattern of FGE variants revealed by native MS (All variants incorporate Cu(I), top row: unlabeled, bottom row: uniformly ^{15}N -labeled). While wild type FGE incorporates 2 Ca(II) ions (left), FGE N188S displays only 1 Ca(II) (middle) and FGE N188I loses both Ca(II) ions due to the disease-related mutations (right). The labeling efficiency of uniformly ^{15}N -labeled variants amounts to 99%.

Table S2: Wild type FGE and disease-related mutants characterized by native MS ($M_{\text{Ca(II)}} = 40.1$ Da, $M_{\text{Cu(I)}} = 63.5$ Da).

Enzyme variant	$M_{\text{W,calc}}$ (Da)	$M_{\text{W,obs}}$ (Da)	Comment	Δm to metal-free state (Da)
FGEwt	33612.7	33608.0	1 Cu(I), 2 Ca(II)	138.6
FGE N188S	33545.6	33543.4	1 Cu(I), 1 Ca(II)	101.2
FGE N188I	33532.7	33531.1	1 Cu(I)	63.1
^{15}N FGEwt	34033.4	34026.6	1 Cu(I), 2 Ca(II)	139.7
^{15}N FGE N188S	33961.8	33961.1	1 Cu(I), 1 Ca(II)	101.8
^{15}N FGE N188I	33953.5	33950.5	1 Cu(I)	64.6

Protein expression and purification

Proteins were expressed and purified as described in Miarzlou et al.¹³

Backbone back-exchange to protonated state after protein expression for the sample with uniformly ²H, ¹³C, ¹⁵N-labeled FGEwt with Cu(I) and two Ca(II)

To perform the backbone amide hydrogen exchange the buffer in the protein sample was exchanged to 20 mM glycine buffer (pH 8.2) with Amicon Ultra centrifugal filters (15 mL, membrane NMWL of 10 kDa). Then the protein sample was diluted to the final concentration of 20 μM in 20 mM glycine buffer (pH 8.2) supplemented with 2 mM CaCl₂, 2 mM DTT and 1 M guanidinium chloride and incubated for 96 h at rt. Afterwards, the buffer was exchanged to 20 mM phosphate buffer (pH 8.2) supplemented with 50 mM NaCl with Amicon Ultra centrifugal filters (15 mL, membrane NMWL of 10 kDa), concentrated, flash-frozen in liquid N₂ and stored at -80 °C.

Sequence

1	11	21	31
MPSFDFDIPR	RSPQEIAKGM	VAIPGGTFRM	GGEDPDAFPE
41	51	61	71
DGEGPVRTVR	LSPFLIDRYA	VSNRQFAAFV	KATGYVTDAE
81	91	101	111
RYGWSFVFHA	HVAPGTPVMD	AVVPEAPWWV	AVPGAYWKAP
121	131	141	151
EGPGSSITDR	PNHPVVHVS	NDAVAYATWA	GKRLPTEAEW
161	171	181	191
EMAARGGLDQ	ARYPWGNELT	PRGRHRCNIW	QGTFPVHDTG
201	211	221	231
EDGYTGAPV	NAFAPNGYGL	YNVAGNVWEW	CADWWSADWH
241	251	261	271
ATESPATRID	PRGPETGTAR	VTKGGNFLCH	ESYCNRYRVA
281	291	301	
ARTCNTPDSS	AAHTGFRCAA	DP	

Native MS. All protein samples were dialyzed in 50 mM (NH₄)₂CO₃ buffer with pH 8.2. For HRMS measurements protein samples were diluted to a concentration of 2 mg/mL in 2 mM DTT and directly injected into HRMS instrument (maXis II ESI-TOF, Bruker). The obtained MS data was deconvoluted using the Bruker Compass DataAnalysis 4.4 SR1 software.

Differential Scanning Fluorimetry (DSF). The proteins were dissolved to the final concentration of 3 μM in 20 mM phosphate buffer (pH 8.0) containing 10 mM NaCl and 1 mM DTT. The samples were loaded into nanoDSF capillaries and subjected to thermal unfolding using the Prometheus NT 48 (NanoTemper Technologies). Thermal unfolding was measured in triplicates with a linear thermal ramp of 1 °C/min from 20 to 95 °C. Unfolding transition points were determined from the change in tryptophan

fluorescence emission wavelengths at 330 and 350 nm, by identifying the maximum of the first derivative of the 350 nm/330 nm ratio against temperature. A linear baseline correction was applied to the data.

NMR experiments

For backbone assignment and all spectra recorded for protein characterization of uniformly ^2H , ^{13}C , ^{15}N labeled FGEwt with 1.5 eq. Ag(I) and 1.1 eq. Cu(I), TROSY³⁶ variants of HSQC, HNC0, HNCA, HN(CO)CA, HN(CA)CO, HNCACB, ^{15}N T_1 and T_2 relaxation and heteronuclear Overhauser effects (heteronuclear NOE) experiments were used.³⁷⁻⁴¹ NMR spectra were recorded on a Bruker Avance III HD spectrometer operating at 600.13 MHz proton frequency equipped with a cryogenic $^1\text{H}/^{19}\text{F}$ - $^{13}\text{C}/^{15}\text{N}$ -D QCI-F probehead. The obtained NMR spectra were assigned using CcpNmr Analysis⁴² and CARA⁴³.

Amide exchange experiments

FGEwt was prepared according to the sample conditions given below in 90% D_2O + 10% H_2O (amide exchange) or 90% H_2O + 10% D_2O (blank). The data points were normalized to an intensity threshold of 20000 at a concentration of 280 μM for an HSQC experiment with 32 scans in order to enable a comparison between the different FGE variants/measurements.

Sample composition FGEwt + 1.5 eq. Cu(I): 275 μM ^{15}N -labeled FGEwt + 1.5 eq. Cu(I), in 2 mM DTT, 20 mM phosphate and 50 mM NaCl with pH 8.2.

Sample composition FGEwt: 280 μM ^{15}N -labeled FGEwt, in 2 mM DTT, 20 mM phosphate and 50 mM NaCl with pH 8.2.

Sample composition FGE N188S + 1.5 eq. Cu(I): 330 μM ^{15}N -labeled FGE N188S + 1.5 eq. Cu(I), in 2 mM DTT, 20 mM phosphate and 50 mM NaCl with pH 8.2.

Sample composition FGE N188I + 1.5 eq. Cu(I): 325 μM ^{15}N -labeled FGE N188I + 1.5 eq. Cu(I), in 2 mM DTT, 20 mM phosphate and 50 mM NaCl with pH 8.2.

Samples for backbone assignment

Sample composition FGEwt + 1.5 eq. Ag(I): 850-1000 μM ^2H , ^{13}C , ^{15}N -labeled FGEwt + 1.5 eq. Ag(I), in 2 mM DTT, 20 mM phosphate and 75 mM NaCl with pH 8.0 measured at a temperature of 310 K. Assignments were transferred to 298 K with the aid of a temperature series.

Sample composition FGEwt + 1.5 eq. Cu(I): 850 μM ^2H , ^{13}C , ^{15}N -labeled FGEwt + 1.5 eq. Cu(I), in 2 mM DTT, 20 mM phosphate and 75 mM NaCl with pH 8.0 measured at a temperature of 310 K. Assignments were transferred to 298 K with the aid of a temperature series.

CSPs outer Ca(II) binding site N188S

Table S3: Small CSPs are found for residues in outer Ca(II) binding site (Comparison ^{15}N FGEwt with Cu(I) and ^{15}N FGE N188S with Cu(I)). Peaks assigned with uniformly $^2\text{H}, ^{13}\text{C}, ^{15}\text{N}$ -labeled samples (FGEwt + Cu(I) with two Ca(II) ions -> assignment for ^{15}N FGEwt with Cu(I), FGEwt + Ag(I) with one Ca(II) ion -> assignment for ^{15}N FGE N188S with Cu(I)).

Residue	Shift ^1H (ppm)	Shift ^{15}N (ppm)
187Cys	0.07	0.66
188Asn	-0.02	0.08
191Gln	0.03	0.02
192Gly	0.00	0.06
199Thr	-0.03	0.00
200Gly	-0.06	-0.39
201Glu	0.00	-0.14
202Asp	0.00	-0.18
203Gly	0.02	0.20
204Tyr	0.02	-0.07

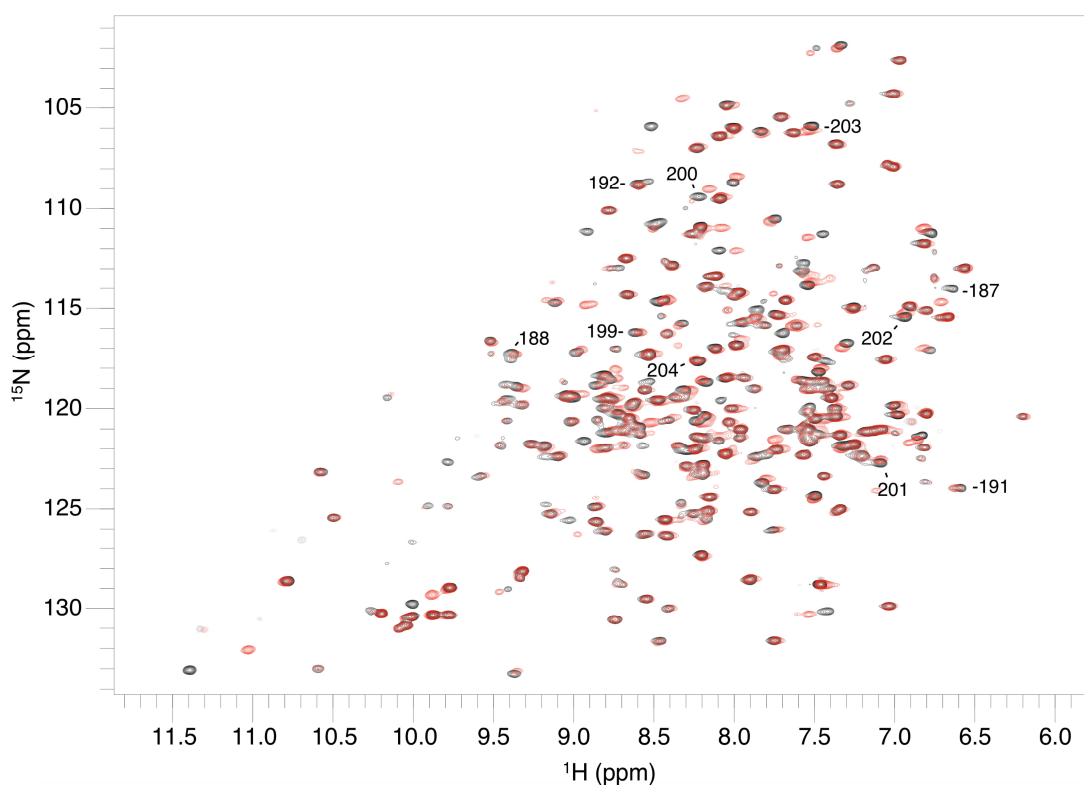


Figure S5: Overlay of ^1H - ^{15}N TROSY-HSQC spectra of ^{15}N FGEwt with Cu(I) (black) and ^{15}N FGE N188S with Cu(I) (red). Small CSPs are found for the residues in the outer Ca(II) binding site.

CSPs inner Ca(II) binding site N188S

Table S4: From the comparison of ^{15}N FGEwt with Cu(I) and ^{15}N FGE N188S with Cu(I) it is visible, that the signals corresponding to the inner Ca(II) binding site exhibit either large CSPs or are not traceable in the ^{15}N FGE N188S with Cu(I) sample. Most close candidate peaks are given in the table below. Peaks for the inner Ca(II) binding site were assigned using the uniformly ^2H , ^{13}C , ^{15}N -labeled FGEwt + Ag(I) with one Ca(II) ion and transferred to ^{15}N FGE N188S with Cu(I).

Residue	Minimal shift ^1H (ppm) (or not traceable)	Minimal shift ^{15}N (ppm) (or not traceable)
222Asn	> 0.3	> 1.0
223Val	> 0.8	> 3.0
224Ala	> 0.3	> 2.0
225Gly	> 0.2	> 1.4
227Val	> 0.1	> 1.5

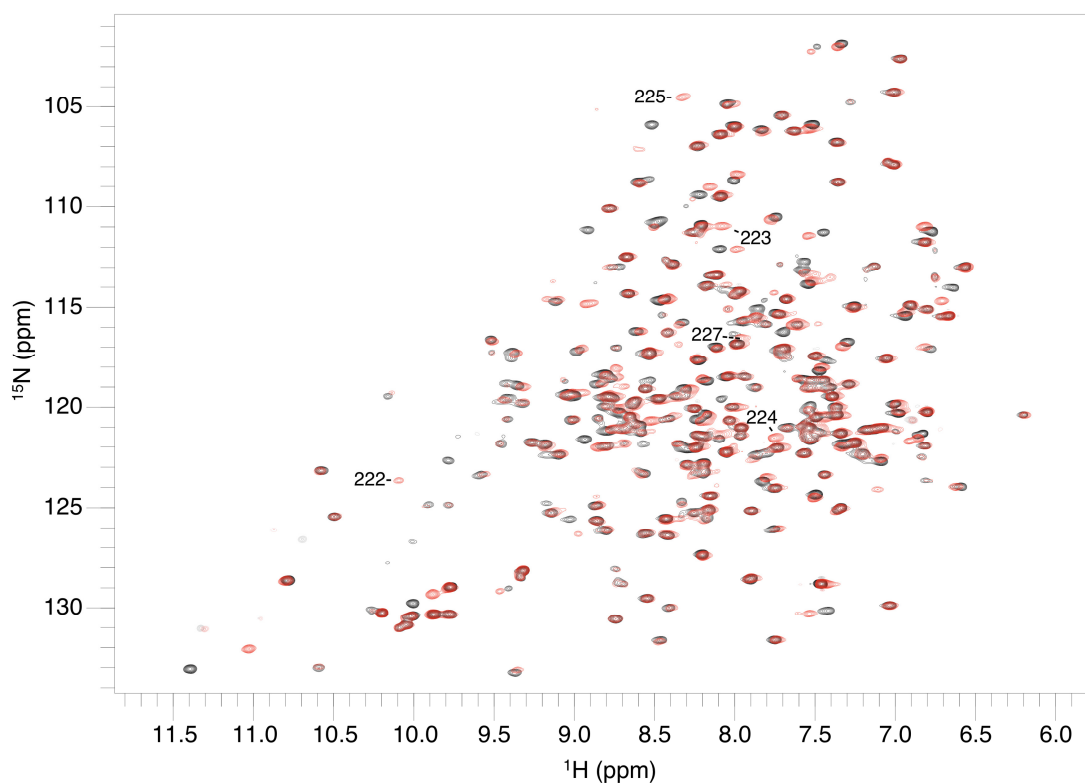


Figure S6: From the comparison of ^{15}N FGEwt with Cu(I) (black) and ^{15}N FGE N188S with Cu(I) (red) it is visible, that the signals corresponding to the inner Ca(II) binding site exhibit either large CSPs or are not traceable in the ^{15}N FGE N188S with Cu(I) sample.

CSPs outer Ca(II) binding site N188I

Table S5: From the comparison of ^{15}N FGEwt with Cu(I) and ^{15}N FGE N188I with Cu(I) it is visible, that the signals corresponding to the outer Ca(II) binding site exhibit either large CSPs or are not traceable in the ^{15}N FGE N188I with Cu(I) sample. Most close candidate peaks are given in the table below. Peaks for the outer Ca(II) binding site were assigned using the uniformly ^2H , ^{13}C , ^{15}N -labeled FGEwt + Ag(I) with one Ca(II) ion as well as ^{15}N FGE N188S with Cu(I) and transferred to ^{15}N FGE N188I with Cu(I).

Residue	Minimal shift ^1H (ppm) (or not traceable)	Minimal shift ^{15}N (ppm) (or not traceable)
187Cys	> 0.3	> 2.5
188Asn	> 0.2	> 1.7
191Gln	> 0.7	> 3.0
192Gly	> 1.2	> 0.0
199Thr	> 0.5	> 3.1
200Gly	> 0.4	> 0.7
201Glu	> 0.4	> 1.0
202Asp	> 0.3	> 2.4
203Gly	> 0.3	> 2.8
204Tyr	> 0.3	> 1.5

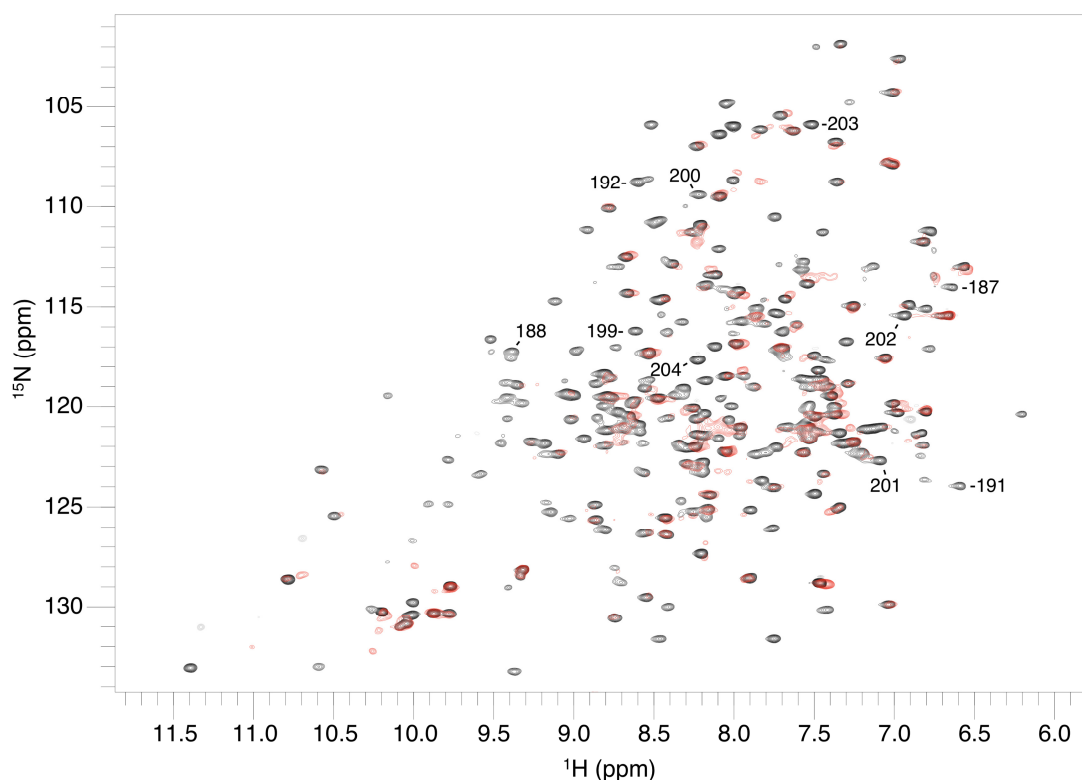


Figure S7: From the comparison of ^{15}N FGEwt with Cu(I) (black) and ^{15}N FGE N188I with Cu(I) (red) it is visible, that all signals corresponding to the outer Ca(II) binding site exhibit either large CSPs or are not traceable in the ^{15}N FGE N188I with Cu(I) sample.

Amide exchange FGEwt + 1.5 eq. Cu(I)

Table S6: Number of peaks over the intensity threshold above 20000 at 280 μ M with 32 scans for the amide exchange of FGEwt + 1.5 eq. Cu(I).

Data point	Time (min)	Number of peaks	Remaining peaks (%)
blank	0	240	100
1	161	93	39
2	974	83	35
3	1370	79	33
4	1907	75	31
5	2699	74	31
6	3527	79	33
7	4319	76	32

Amide exchange FGEwt

Table S7: Number of peaks over the intensity threshold above 20000 at 280 μ M with 32 scans for the amide exchange of FGEwt.

Data point	Time (min)	Number of peaks	Remaining peaks (%)
blank	0	239	100
1	173	89	37
2	976	78	33
3	1373	77	32
4	1896	66	28
5	2688	64	27
6	3479	63	26
7	4271	63	26

Amide exchange FGE N188S + 1.5 eq. Cu(I)

Table S8: Number of peaks over the intensity threshold above 20000 at 280 μ M with 32 scans for the amide exchange of FGE N188S + 1.5 eq. Cu(I).

Data point	Time (min)	Number of peaks	Remaining peaks (%)
blank	0	240	100
1	151	51	21
2	942	54	23
3	1338	57	24
4	1915	48	20
5	2706	50	21
6	3498	51	21
7	4290	51	21

Amide exchange FGE N188I + 1.5 eq. Cu(I)

Table S9: Number of peaks over the intensity threshold above 20000 at 280 μ M with 32 scans for the amide exchange of FGE N188I + 1.5 eq. Cu(I).

Data point	Time (min)	Number of peaks	Remaining peaks (%)
blank	0	165	100
1	138	4	2
2	932	5	3
3	1328	4	2
4	1905	5	3
5	2696	4	2
6	3487	4	2
7	4279	5	3

Amide exchange: Remaining peaks grouped in three different regimes (fast, medium, slow)

Table S10: Remaining peaks in amide exchange experiments sorted in three different groups: fast exchange (< 3 h), medium exchange (< 1 d) and slow exchange (> 1 d).

Regime	Time	Number of peaks	Number of peaks (%)	FGE variant
fast	< 3 h	147	61	FGEwt + 1.5 eq. Cu(I)
medium	< 1 d	14	6	
slow	> 1 d	79	33	
fast	< 3 h	150	63	FGEwt
medium	< 1 d	12	5	
slow	> 1 d	77	32	
fast	< 3 h	189	79	FGE N188S + 1.5 eq. Cu(I)
medium	< 1 d	0	0	
slow	> 1 d	51	21	
fast	< 3 h	161	98	FGE N188I + 1.5 eq. Cu(I)
medium	< 1 d	0	0	
slow	> 1 d	4	2	

Kinetic data / rate of product formation

Table S11: Rates of product formation catalyzed by different FGE variants. The activity was measured under standard conditions in presence of 5 mM DTT, 1.0 eq. Cu(I) and 200 μ M FGE-27 (peptide substrate).

FGE variant	Rate (1/min)
FGEwt + 1.0 eq. Cu(I)	2.72
FGE N188S + 1.0 eq. Cu(I)	2.37
FGE N188I + 1.0 eq. Cu(I)	1.81

Table S12: Sequence of FGE-27, a 17 amino acid long peptide substrate based on the amino acid sequence of a sulfatase and designed for kinetic measurements of FGE variants.⁴⁴

Residue number	Amino acid
1	Abz-A
2	T
3	T
4	P
5	L
6	C
7	G
8	P
9	S
10	R
11	A
12	S
13	I
14	L
15	S
16	G
17	R-NH ₂

¹⁵N T₁, T₂ relaxation and heteronuclear NOE measured on uniformly ²H,¹³C,¹⁵N-labeled FGEwt with 1.5 eq. Ag(I)

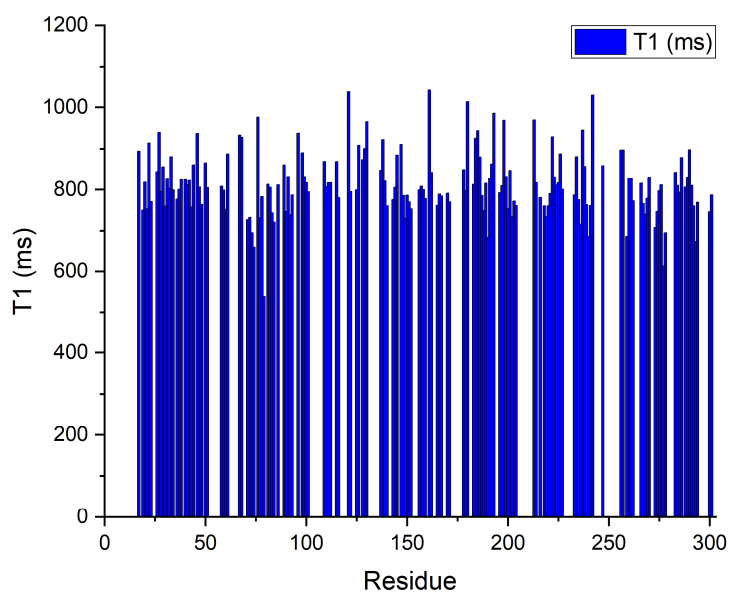


Figure S8: ¹⁵N T₁ relaxation times for uniformly ²H,¹³C,¹⁵N-labeled FGEwt with 1.5 eq. Ag(I).

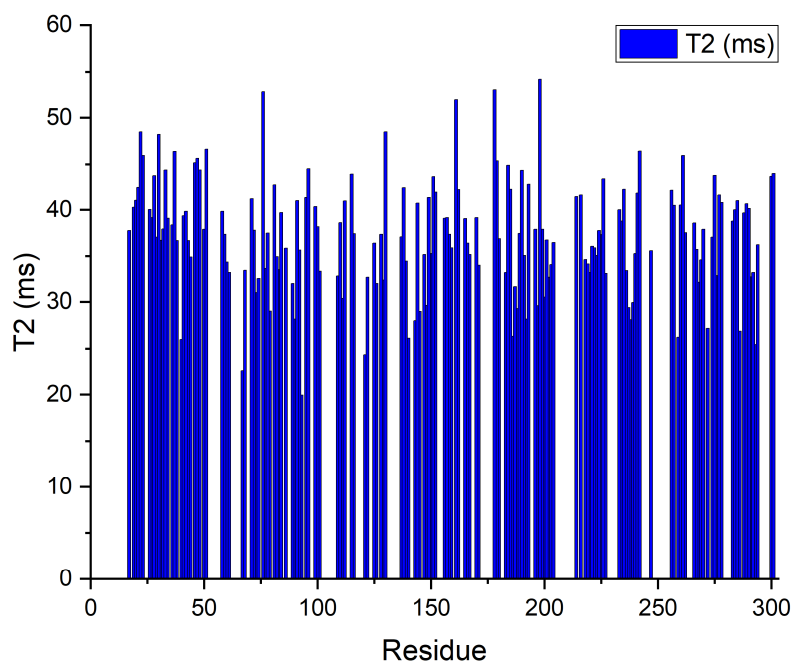


Figure S9: ¹⁵N T₂ relaxation times for uniformly ²H,¹³C,¹⁵N-labeled FGEwt with 1.5 eq. Ag(I).

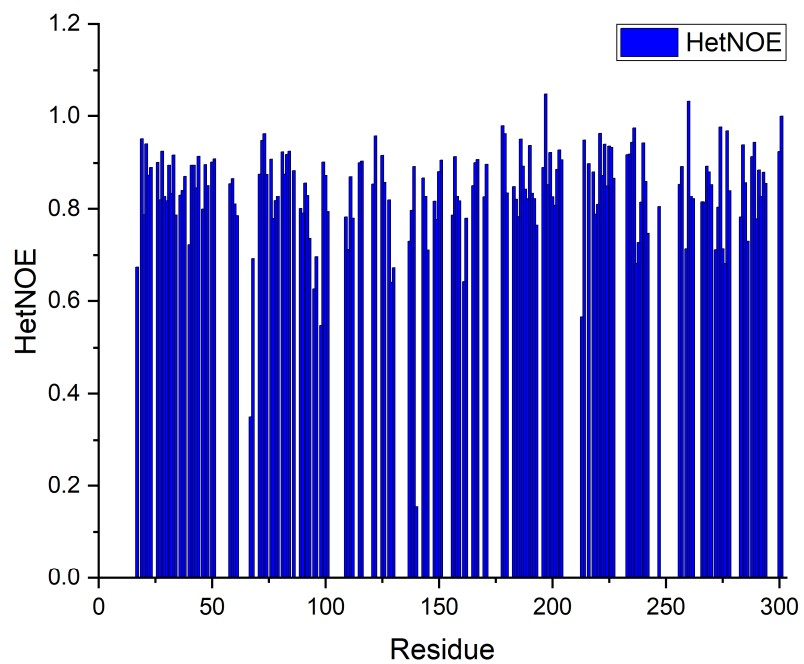


Figure S10: Heteronuclear NOEs for uniformly ^2H , ^{13}C , ^{15}N -labeled FGEwt with 1.5 eq. Ag(I).

¹⁵N T₁, T₂ relaxation and heteronuclear NOE measured on uniformly ²H,¹³C,¹⁵N-labeled FGEwt with 1.1 eq. Cu(I)

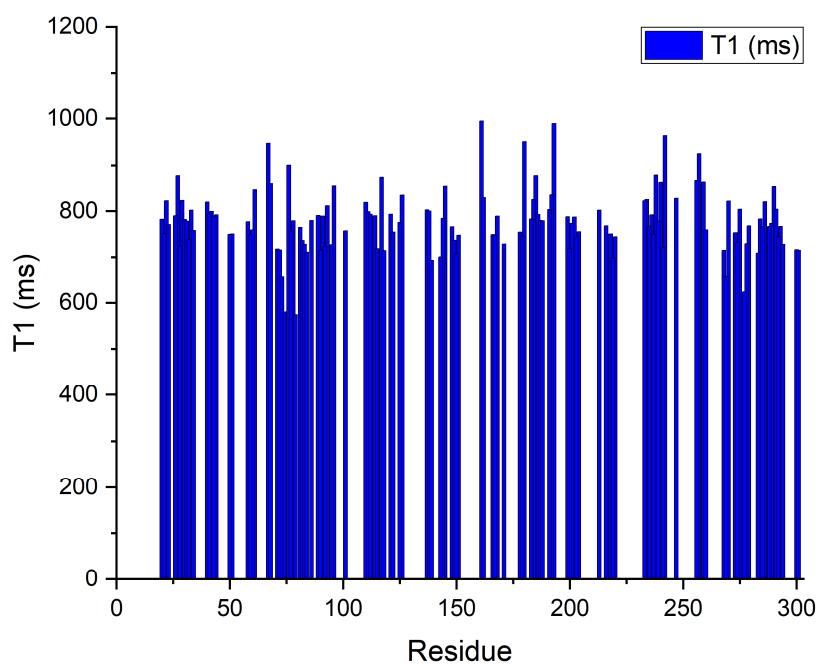


Figure S11: ¹⁵N T₁ relaxation times for uniformly ²H,¹³C,¹⁵N-labeled FGEwt with 1.1 eq. Cu(I).

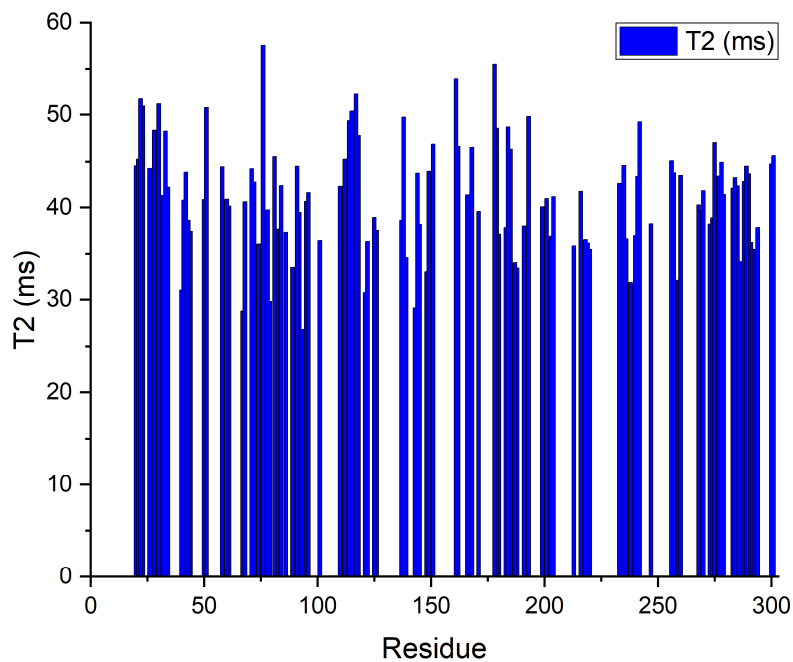


Figure S12: ¹⁵N T₂ relaxation times for uniformly ²H,¹³C,¹⁵N-labeled FGEwt with 1.1 eq. Cu(I).

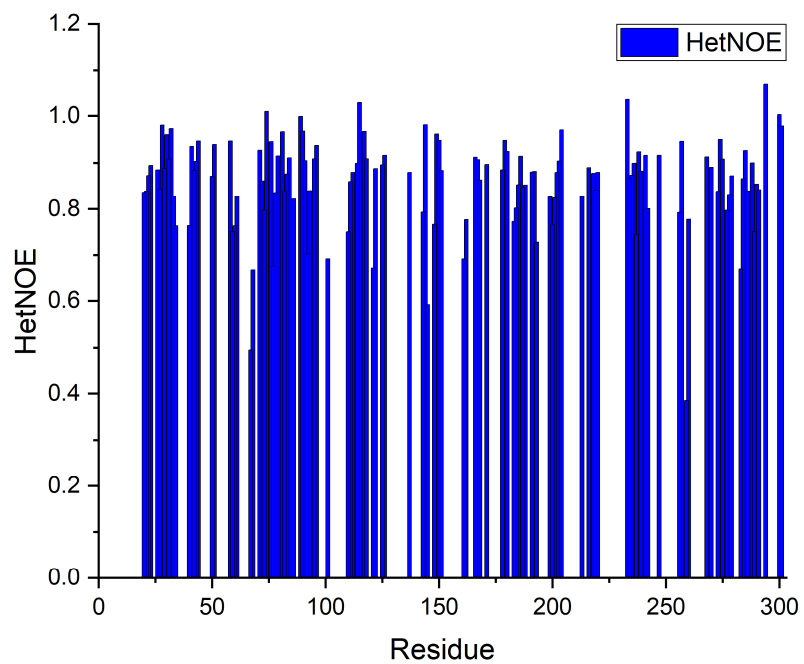


Figure S13: Heteronuclear NOEs for uniformly ^2H , ^{13}C , ^{15}N -labeled FGEwt with 1.1 eq. Cu(I).

References

1. B. Alberts, Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P., *Molecular Biology of the Cell*, 4th edition. Garland Science: New York, 2002.
2. L. Pauling, *Nature*, **1948**, *161*, 707-709.
3. L. Pauling, *Chem. Eng. News*, **1946**, *24*, 1375-1377.
4. L. Cao, O. Caldararu, A. C. Rosenzweig, U. Ryde, *Angew. Chem. Int. Ed.*, **2018**, *57*, 162-166.
5. K. E. H. Frandsen, T. J. Simmons, P. Dupree, J.-C. N. Poulsen, G. R. Hemsworth, L. Ciano, E. M. Johnston, M. Tovborg, K. S. Johansen, P. von Freiesleben, L. Marmuse, S. Fort, S. Cottaz, H. Driguez, B. Henrissat, N. Lenfant, F. Tuna, A. Baldansuren, G. J. Davies, L. Lo Leggio, P. H. Walton, *Nat. Chem. Biol.*, **2016**, *12*, 298-303.
6. S. T. Prigge, B. A. Eipper, R. E. Mains, L. M. Amzel, *Science*, **2004**, *304*, 864.
7. S. T. Prigge, A. S. Kolhekar, B. A. Eipper, R. E. Mains, L. M. Amzel, *Nat. Struct. Biol.*, **1999**, *6*, 976-983.
8. R. J. Quinlan, M. D. Sweeney, L. Lo Leggio, H. Otten, J.-C. N. Poulsen, K. S. Johansen, K. B. R. M. Krogh, C. I. Jørgensen, M. Tovborg, A. Anthonsen, T. Tryfona, C. P. Walter, P. Dupree, F. Xu, G. J. Davies, P. H. Walton, *Proc. Natl. Acad. Sci.*, **2011**, *108*, 15079-15084.
9. T. V. Vendelboe, P. Harris, Y. Zhao, T. S. Walter, K. Harlos, K. El Omari, H. E. M. Christensen, *Sci. Adv.*, **2016**, *2*, e1500980.
10. C. M. Wilmot, J. Hajdu, M. J. McPherson, P. F. Knowles, S. E. V. Phillips, *Science*, **1999**, *286*, 1724-1728.
11. A. Changela, K. Chen, Y. Xue, J. Holschen, C. E. Outten, T. V. Halloran, A. Mondragón, *Science*, **2003**, *301*, 1383.
12. A. K. Wernimont, D. L. Huffman, A. L. Lamb, T. V. O'Halloran, A. C. Rosenzweig, *Nat. Struct. Biol.*, **2000**, *7*, 766-771.
13. D. A. Miarzlou, F. Leisinger, D. Joss, D. Häussinger, F. P. Seebeck, *Chem. Sci.*, **2019**, *10*, 7049-7058.
14. M. Knop, P. Engi, R. Lemnar, F. P. Seebeck, *ChemBioChem*, **2015**, *16*, 2147-2150.
15. M. Knop, T. Q. Dang, G. Jeschke, F. P. Seebeck, *ChemBioChem*, **2017**, *18*, 161-165.
16. M. Meury, M. Knop, F. P. Seebeck, *Angew. Chem. Int. Ed.*, **2017**, *56*, 8115-8119.
17. J. H. Austin, *JAMA Neurol.*, **1973**, *28*, 258-264.
18. G. Diez-Roux, A. Ballabio, *Annu. Rev. Genomics Hum. Genet.*, **2005**, *6*, 355-379.
19. M. J. Appel, C. R. Bertozzi, *ACS Chem. Biol.*, **2015**, *10*, 72-84.
20. P. Bojarová, S. J. Williams, *Curr. Opin. Chem. Biol.*, **2008**, *12*, 573-581.
21. F. Leisinger, D. A. Miarzlou, F. P. Seebeck, *Angew. Chem. Int. Ed.*, **2021**, *60*, 6154-6159.
22. F. Sabourdy, L. Mourey, E. Le Trionnaire, N. Bednarek, C. Caillaud, Y. Chaix, M.-A. Delrue, A. Dusser, R. Froissart, R. Garnotel, N. Guffon, A. Megarbane, H. Ogier de Baulny, J.-M. Pédespan, S. Pichard, V. Valayannopoulos, A. Verloes, T. Levade, *Orphanet J. Rare Dis.*, **2015**, *10*, 31.
23. M. P. Cosma, S. Pepe, G. Parenti, C. Settembre, I. Annunziata, R. Wade-Martins, C. D. Domenico, P. D. Natale, A. Mankad, B. Cox, G. Uziel, G. M. S. Mancini, E. Zammarchi, M. A. Donati, W. J. Kleijer, M. Filocamo, R. Carozzo, M. Carella, A. Ballabio, *Hum. Mutat.*, **2004**, *23*, 576-581.
24. G. N. Ramachandran, C. Ramakrishnan, V. Sasisekharan, *J. Mol. Biol.*, **1963**, *7*, 95-99.
25. J. Wuerges, G. Garau, S. Geremia, S. N. Fedosov, T. E. Petersen, L. Randaccio, *Proc. Natl. Acad. Sci.*, **2006**, *103*, 4386-4391.
26. R. Liddington, Z. Derewenda, E. Dodson, R. Hubbard, G. Dodson, *J. Mol. Biol.*, **1992**, *228*, 551-579.
27. H. K. Song, S. W. Suh, *J. Mol. Biol.*, **1998**, *275*, 347-363.
28. M. Heinig, D. Frishman, *Nucleic Acids Res.*, **2004**, *32*, W500-W502.
29. L. Schlotawa, K. Radhakrishnan, M. Baumgartner, R. Schmid, B. Schmidt, T. Dierks, J. Gärtner, *Eur. J. Hum. Genet.*, **2013**, *21*, 1020-1023.
30. T. Dierks, A. Dickmanns, A. Preusser-Kunze, B. Schmidt, M. Mariappan, K. von Figura, R. Ficner, M. G. Rudolph, *Cell*, **2005**, *121*, 541-552.
31. Z. Yuan, F. Gao, G. Bai, H. Xia, L. Gu, S. Xu, *Biochem. Biophys. Res. Commun.*, **2017**, *484*, 195-201.
32. R. Jain, J. P. Shapleigh, *Microbiology (Reading, Engl.)*, **2001**, *147*, 2505-2515.
33. A. R. Stampfli, K. V. Goncharenko, M. Meury, B. N. Dubey, T. Schirmer, F. P. Seebeck, *J. Am. Chem. Soc.*, **2019**, *141*, 5275-5285.
34. K. V. Goncharenko, A. Vit, W. Blankenfeldt, F. P. Seebeck, *Angew. Chem. Int. Ed.*, **2015**, *54*, 2821-2824.
35. E. M. Tichy, S. W. Hardwick, B. F. Luisi, G. P. C. Salmond, *Acta Crystallogr. D*, **2017**, *73*, 549-556.

36. K. Pervushin, R. Riek, G. Wider, K. Wuthrich, *Proc. Natl. Acad. Sci.*, **1997**, *94*, 12366-12371.
37. M. Salzmann, K. Pervushin, G. Wider, H. Senn, K. Wuthrich, *Proc. Natl. Acad. Sci.*, **1998**, *95*, 13585-13590.
38. M. Salzmann, G. Wider, K. Pervushin, K. Wuthrich, *J. Biomol. NMR*, **1999**, *15*, 181-184.
39. M. Salzmann, G. Wider, K. Pervushin, H. Senn, K. Wuthrich, *J. Am. Chem. Soc.*, **1999**, *121*, 844-848.
40. G. Zhu, Y. Xia, L. K. Nicholson, K. H. Sze, *J. Magn. Reson.*, **2000**, *143*, 423-426.
41. C. Renner, M. Schleicher, L. Moroder, T. A. Holak, *J. Biomol. NMR*, **2002**, *23*, 23-33.
42. W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides, E. D. Laue, *Proteins*, **2005**, *59*, 687-696.
43. R. L. J. Keller, *ETH Zurich*, **2004**, *PhD Thesis (ETH) Nr. 15947*.
44. T. Barbeyron, L. Brillet-Guéguen, W. Carré, C. Carrière, C. Caron, M. Czjzek, M. Hoebeke, G. Michel, *PLoS One*, **2016**, *11*, e0164846.

7 Part V

7.1 T cell activation by nucleobase adduct-containing metabolites

This project was performed in collaboration with the research group of Prof. Dr. Gennaro De Libero from the University of Basel.

T cell activation by nucleobase adduct-containing metabolites

Alessandro Vacchini^{1,†}, Andrew Chancellor^{1,†}, Qinmei Yang^{1,†}, Julian Spagnuolo^{1,†}, Daniel
Joss², Ove Øyås^{3,4}, Aisha Beshirova¹, Corinne De Gregorio¹, Michael Pfeffer², Alfred
5 Zippelius⁵, Jörg Stelling³, Daniel Häussinger², Marco Lepore^{1,6}, Lucia Mori¹, Gennaro De
Libero^{1*}

¹Experimental Immunology, Department of Biomedicine, University Hospital and University of
Basel, 4031 Basel, Switzerland

10 ²Department of Chemistry, University of Basel, 4056 Basel, Switzerland

³Department of Biosystems Science and Engineering and SIB Swiss Institute of Bioinformatics,
ETH Zurich, 4058 Basel, Switzerland

⁴Present address: Department of Animal and Aquacultural Sciences, Norwegian University of
Life Sciences, 1432 Ås, Norway

15 ⁵Laboratory of Cancer Immunology, Department of Biomedicine, University of Basel and
University Hospital of Basel, 4031 Basel, Switzerland

⁶Present address: Immunocore Ltd, Abingdon, Oxfordshire, OX14 4RX, United Kingdom

[†]These authors contributed equally

20 *Lead Contact. Correspondence: gennaro.delibero@unibas.ch (G.D.L.)

One Sentence Summary

Reactive carbonyls accumulating in metabolically altered cells modify purines and pyrimidines
to generate novel T cell antigens.

25

Abstract:

T lymphocytes recognize antigens derived from a variety of chemical structures. Here, we describe a novel class of T cell-stimulatory molecules made of self-DNA and RNA adducts, presented by the non-polymorphic MHC-class-I-related molecule MR1. The nucleobase adducts accumulate upon combined alterations in the metabolism of nucleobases, carbonyls and reactive oxygen species and stimulate metabolite-specific T cells (MR1T cells). The metabolic changes generating nucleobase adducts preferentially occur in tumor cells independently of tissue origin, explaining MR1T cell recognition of many different cancer types. This novel type of antigen specificity suggests that MR1T cells survey cells for metabolic alterations, laying the groundwork for novel genetically-unrestricted T cell-based immunotherapies targeting the altered metabolome.

MR1T lymphocytes are a recently discovered T cell population that recognize non-microbial cell-endogenous antigens presented by the monomorphic MHC class I-related molecule MR1, use a polyclonal T-cell receptor (TCR) repertoire, and exhibit diverse T-helper-like functions (1). Studies have since revealed that these MR1T cells can efficiently recognize and kill various tumor cell types (1, 2), thus representing promising candidates for cancer immunotherapy (3). However, the molecular basis for differential recognition of cancer cells versus healthy tissue and identification of MR1T cell ligands is still unknown.

MR1 is a ubiquitously expressed molecule that was initially characterized as the restriction element for mucosal-associated invariant-T (MAIT) cells (4), enabling MAIT cell detection of microbes upon infection (5-7) by recognition of metabolites produced in the riboflavin pathway (7, 8). Subsequent studies revealed that the MR1 antigen-binding pocket can accommodate various metabolites and chemical entities (7-11). Accordingly, the MR1T cell population also seems highly heterogeneous (1, 2, 12). The challenge now is to understand which antigens are key for stimulating specific members of the MR1T cell population under different physiological and pathological conditions.

In the case of tumor cells, we hypothesized that the MR1T-cell antigens might be tumor cell-specific metabolites. Tumor cells undergo dramatic metabolic changes that confer a selective advantage and promote their survival, proliferation and migration (13, 14). Wide gene deletion technology to disrupt tumor cell metabolism disclosed the processes that are key for MR1T-cell recognition. In complementary studies, we identified MR1T-cell stimulatory metabolites and generated MR1-metabolite tetramers that stained MR1T cells in tumor-infiltrating lymphocytes (TILs). Our results explain observations of preferential tumor cell recognition by MR1T cells, highlighting their potential as surveyors of cellular metabolic integrity and possible application for genetically-unrestricted T cell therapies targeting the tumor metabolome.

Results

65

Knock-out of single metabolic pathway genes within target cells impacts MR1T-cell recognition

We first used a genome-wide CRISPR/Cas9-mediated gene disruption approach (15) to identify the cellular pathways involved in tumor cell susceptibility to MR1T-cell killing.

70 We hypothesized that inactivation of genes involved in antigen biosynthesis would positively or negatively affect accumulation of antigenic compounds, thereby modifying MR1T-cell recognition of tumor cells.

We transduced A375 melanoma tumor cells stably engineered with the MR1 and CAS9 genes (A375-MR1 cells) with a library of single guide RNAs (sgRNAs) covering the total human
75 genome to generate targets for the cytotoxic MR1T cell clone TC5A87. After sequential rounds of T-cell killing, each followed by a recovery phase, surviving cells were analyzed by amplicon-sequencing for enriched or depleted sgRNAs compared to the same cells not subjected to killing. We removed genes essential for A375-MR1 cell growth (16) from the dataset to limit the impact of T-cell-independent processes. We defined significantly enriched or depleted
80 sgRNAs (FDR>0.05) based on a log₂ fold-change cut-off relative to the top and bottom 1% in the control (Fig. 1A). Our approach identified 243 sgRNAs targeting 237 unique genes that were enriched, including MR1 and beta-2-microglobulin (B2M), as expected, which are both essential for MR1-dependent antigen presentation (Fig. 1A). We also found enrichment of sgRNAs specific for the adhesion molecules CD58 (LFA-3) and ICAM-1 (Data S1), which bind
85 the T-cell surface molecules CD2 and CD11a, respectively, and have a fundamental role in mediating target cell recognition and killing (17). Furthermore, we identified significant depletion of 5,331 sgRNAs targeting 4,705 unique genes (Fig. 1A). A far greater number of depletions compared to enrichments was expected, as the number of genes required for tumor cell survival in general is likely to outweigh the number of genes that would allow specific escape from T-cell-mediated killing. Despite this fact, when we applied binomial
90 enrichment analysis of gene-ontology (GO) terms (18, 19) annotated to significant hits, we revealed a large number of both enriched and depleted gene targets (n=127 and 650, respectively), associated with Metabolic Processes (Fig. 1B, C), suggesting that the coordinated activity of multiple genes involved in target cell metabolism in general promoted MR1T-cell recognition. These hits showed enrichment in Nucleobase and Nucleic Acid
95 Metabolic Processes, arguing for a potential role for these metabolic pathways in enhancing MR1T-cell stimulation.

To explore the potential roles of the identified gene targets we used Recon3D, a genome-scale *in silico* model of human metabolism (20). Specifically, we applied structural sensitivity
100 analysis (21, 22) to predict the metabolic network response to single-gene knockouts (KOs). High Pearson correlations (score > 0.6) identified sets of two or more gene KOs predicted to have similar global effects on metabolic reactions (Fig. 1D) and highlighted 125 genes among those identified by the CRISPR/Cas9 screening. Annotation of these genes with their KEGG

105 pathways revealed that oxidative phosphorylation and purine metabolism were affected by single-gene knockouts of a number of different gene targets, leading to common metabolic responses (Fig. 1D).

Taken together, our analyses following a CRISPR/Cas9 screen outlined the importance of several metabolic pathways for tumor recognition by MR1T cells.

110 **Nucleobase and nucleoside antigens presented by tumor cells can stimulate MR1T cells**

We focused first on the purine pathway. Here we identified several genes that might be involved in MR1T-cell antigen accumulation (Fig. S7, Data S1-S2): Adenosine Deaminase (*ADA*), Adenylosuccinate Synthase 1 (*ADSSL1*), Laccase Domain Containing 1 (*LACC1*), cGMP-specific 3',5'-cyclic phosphodiesterase (*PDE5A*), Aldehyde Dehydrogenase 16 Family Member A1 (*ALDH16A1*) and Hypoxanthine Phosphoribosyltransferase 1 (*HPRT1*). All these genes contribute to different activities necessary for the *de novo* purine synthesis and the purine salvage pathway (23-25), supporting purines as molecules potentially involved in target recognition by MR1T cells.

To validate the relevance of the selected genes, we measured the ability of individual KO cell lines to stimulate IFN- γ production by two different MR1T cell clones. *ADA*- and *LACC1*-deficient A375-MR1 cells induced significantly more IFN- γ production by both MR1T cell clones than did the parental A375-MR1 cells (Fig. 2A, B and 2D, E), whereas the stimulatory ability of *ADSSL1*-deficient cells was significantly reduced, as predicted by the CRISPR/Cas9 screening (Fig. 2C and 2F). The gene KO did not impact on MR1 cell surface levels or 5-OP-RU presentation to MR1-restricted MAIT cells (Fig. S1), thus excluding general alterations to the MR1 antigen-presentation pathway. These findings confirm that purines could be potential MR1T-cell antigens.

To investigate the possible role of purines in MR1T-cell stimulation, we incubated human acute monocytic leukemia THP-1 cells with synthetic nucleotides, nucleosides or nucleobases before adding MR1T-cell clones and measuring IFN- γ production. THP-1 cells were selected as targets because they constitutively express low surface levels of MR1 and induce some spontaneous MR1T-cell stimulation, demonstrating their ability to appropriately process and present MR1T-cell antigens. A375-MR1 cells expressing very high surface levels of MR1 were also included as a positive control to stimulate MR1T-cell cytokine production. Three MR1T-cell clones differently reacted to tested compounds: TC5A87 did not show significant reactivity (Fig. 2G, left); DGB129 reacted to adenine, adenosine, deoxyadenosine (dAdo) and inosine (Fig. 2G, middle); and MCA3C3 was activated by ADP, guanine, guanosine, deoxyguanosine and xanthosine (Fig. 2G, right). THP-1 cells incubated with the synthetic compounds did not stimulate MAIT cells (Fig. S2A). Interestingly, despite the compounds were used at high concentration, their stimulatory effect was minimal compared to MR1T stimulation with A375-MR1, suggesting that these molecules might be intermediate precursors of antigens.

Taken together, these findings indicate that human MR1T cell clones can recognize nucleobase/nucleoside antigens processed by and presented on a cancer cell line, and that cross-reactivity might exist between different MR1T-cell clones.

145 **Methylglyoxal and purine metabolism pathways within APCs cooperate for MR1T-cell stimulation**

To understand which metabolic pathways might be involved in recognizing nucleobase/nucleoside antigens on tumor cells by MR1T cells we began by interrogating our whole-genome gene disruption screening data. We observed that some of the significantly
150 depleted sgRNAs were related to genes involved in glycolysis (*TPI1*) and methylglyoxal (MG) degradation, including Glyoxalase 1 (*GLO1*), and Glyoxalase Domain Containing 4 (*GLOD4*). *TPI1* encodes a triosephosphate isomerase within the glycolytic pathway and is responsible for the enzymatic conversion of dihydroxyacetone phosphate (DHAP) into glyceraldehyde 3-phosphate (G3P) — a reaction that can otherwise occur spontaneously with the generation of
155 MG (26) (Fig. 3). Conversely, *GLO1*-deficient cells are impaired in MG (a highly reactive carbonyl) degradation, which therefore accumulates (27). As MG forms adducts with several nucleobases (28), these data suggest a potential involvement of MG in generating MR1T-cell antigens.

We then dissected the possible roles of glycolysis and MG degradation in MR1T stimulation
160 by generating single gene KO cell lines. Loss of *TPI1* in A375-MR1 cells significantly increased IFN- γ production by both MR1T-cell clones (Fig. 3A, B). Furthermore, A375-MR1 cells pulsed with glucose and then fixed showed increased stimulatory capacity (Fig. 3C, D); this effect was abolished when the same cells were incubated with deoxyglucose (Fig. 3C, D), which does not enter the glycolytic pathway and so does not generate MG (29). We also saw that *GLO1*-
165 deficient and *GLO1*-overexpressing A375-MR1 cells showed increased and reduced stimulatory capacity, respectively (Fig. 3E, F). Altogether these results suggest that MG accumulation in target cells is important for the stimulation of MR1T-cell clones.

To investigate the potential synergism between MG and purine metabolic pathways in MR1T-cell stimulatory capacity of tumor cells, we explored the effects of individual and
170 combined pharmacologic inhibition of key enzymes in these pathways. We used *S-p*-bromobenzylglutathione (BBG) to inhibit *GLO1* (30, 31); mycophenolic acid (MPA) to inhibit inosine monophosphate dehydrogenases (*IMPDH1*, and *IMPDH2*) (32), leading to IMP accumulation; and erythro-9-(2-Hydroxy-3-nonyl) adenine hydrochloride (EHNA) to inhibit ADA, inducing adenosine, dAdo and cGMP accumulation (33). To sensitively detect the effects
175 of the inhibitors, we again used THP-1 cells as target cells: we found that BBG in combination

with each of the other two drugs significantly enhanced the IFN- γ release of both MR1T cell clones to THP-1 cells (Fig. 3G, H). The DGB129 clone was more sensitive to these treatments and also reacted to THP-1 cells treated with EHNA, BBG, or MPA alone (Fig. 3H).

180 Finally, we tested the IFN- γ response of MR1T cells to GLO1-modified THP-1 cells and various doses of MG or dAdo. We found that MG treatment significantly increased stimulation by GLO1-deficient compared to wild-type THP-1 cells (Fig. 3I). Conversely, MG failed to induce stimulation when administered to GLO1-overexpressing cells (Fig. 3J). Similarly, dAdo increased the MR1T response to GLO1-deficient THP-1 (Fig. 3K) and decreased MR1T reactivity to GLO1-overexpressing THP-1 cells (Fig. 3L). Together, these findings suggest that
185 nucleosides/nucleobases and MG cooperate in generating potential MR1T-cell antigens.

Multiple oxidative stress-related carbonyl species accumulate within tumor cells and contribute to MR1T cell stimulation

In addition to the purine pathway, our model-based analysis (Fig. 1D) highlighted genes
190 related to oxidative phosphorylation, whose protein products participate in ATP generation within mitochondria and whose alteration promotes accumulation of reactive oxygen species (ROS) (34-36). Therefore, we next investigated the roles of these molecules in MR1T-cell stimulation by tumor cells.

First, we focused on genes involved in oxidative phosphorylation. Our initial MR1T-cell
195 killing screen uncovered a significant depletion of sgRNAs specific for *GSTM1*, *GSTA4*, *GSTA1*, *GSTM5*, *GSTA2*, *GSTM3*, and *GSTO1* (Data S2); these genes are involved in the detoxification of electrophilic compounds and ROS by their conjugation to glutathione (GSH), a ROS scavenger (37, 38). We therefore hypothesized that in the absence of GSTs, and upon accumulation of ROS and electrophilic molecules (39) (Fig. S8), tumor cells may accumulate
200 MR1T-cell stimulatory compounds. Accordingly, we tested the effects of paclitaxel and doxorubicin, two drugs that induce cellular accumulation of O₂⁻ and H₂O₂ (40-42). Both drugs significantly increased ROS accumulation (Fig. S3A) and promoted activation of all three MR1T-cell clones when incubated with THP-1 cells and nucleoside compounds (Fig. 4A-C), mirroring the additive effects observed when purine-modifying drugs and carbonyl-
205 degradation inhibitors were combined (Fig. 3G, H).

Next, we treated A375-MR1 with apocynin, an O₂⁻ scavenger and NADPH oxidase inhibitor (43), or with GSH or N-acetylcysteine (NAC), which prevent H₂O₂ accumulation (40) before fixation and incubation with the three MR1T cell clones. We found that A375-MR1 cells treated with any of the inhibitors stimulated significantly less IFN- γ production from MR1T

210 cells, with apocynin being effective with one T-cell clone (Fig. 4D-F). We also treated A375-
MR1 cells with buthionine sulfoximine (BSO), an inhibitor of GSH synthase (44), and a
significant increase was observed in the stimulation of all the tested MR1T clones (Fig. 4D-F).
Together, these data show that ROS participate in MR1T antigen accumulation, although
requires concomitant alteration of nucleobases metabolism.

215 Peroxide accumulates in many tumor types and is involved in various signal transduction
pathways and cell fate decisions (45). Peroxide is also necessary for lipid peroxidation, a
pathway that generates malondialdehyde (MDA) and 4-OH-nonenal (4-HNE), two highly
reactive carbonyls (46, 47). Both compounds form stable adducts with proteins, lipids and
nucleobases and accumulate within tumor cells (48-52). Alongside our finding that inhibiting
220 ROS accumulation impedes tumor cell stimulation of MR1T cells (Fig. 4D-F), we inferred a role
for lipid peroxidation from the results of our CRISPR/Cas9 screen, which showed significant
depletion of the glutathione peroxidase 4 (GPX4) and glutathione peroxidase 1 (GPX1) sgRNAs
(Fig. 1 and Data S2). While GPX1 protein catalyzes the reduction of organic hydroperoxides
and H₂O₂ by glutathione, GPX4 has a high preference for lipid hydroperoxides (Fig. S8) and
225 protects cells against membrane lipid peroxidation and death (53). Accordingly, when we pre-
treated A375-MR1 cells with the selective GPX1 inhibitor mercaptosuccinic acid (MSA), or with
two GPX4 inhibitors RSL3 and ML-210, they showed significantly increased MR1T-cell
stimulatory activity (Fig. 4G-I). None of these compounds influenced the MAIT-cell response
to microbial antigens in control experiments, except paclitaxel that in presence of nucleosides
230 could induce a little but significant stimulation (Fig. S2C-E). Taken together, these findings
indicate that peroxides and lipid peroxidation contribute to MR1T-cell stimulation by tumor
cells.

To further assess the carbonyl involvement in the generation of MR1T antigens, we tested
the capacity of carbonyl scavengers to prevent MR1T-cell activation. A375-MR1 cells were
235 incubated with aminoguanidine and hydralazine, which show preferential scavenging activity
for different carbonyls (54), then fixed and washed before the addition of MR1T cell clones.
Both scavengers significantly inhibited IFN- γ production by MR1T cell clones (Fig. 4J-L).

Collectively, our data suggest that multiple oxidative stress-related reactive carbonyl
species accumulating in cells following metabolic alterations combine with nucleobases to
240 generate MR1-presented antigens that stimulate MR1T cells.

Nucleobase adducts stimulate MR1T cells

The biochemical condensation of carbonyl species with nucleobases frequently occurs in many cancer cell-types, and leads to the generation of different types of adducts (55). We considered the possibility that nucleobase-adduct-containing compounds might be MR1T-cell antigens. To test this hypothesis, we synthesized four previously described adducts: the purine adducts 8-(9*H*-purin-6-yl)-2-oxa-8-azabicyclo[3.3.1]nona-3,6-diene-4,6-dicarbaldehyde (M₃ADE) (56), N⁶-(3-oxo-1-propenyl)-2'-deoxyadenosine (OPdA) (56), and pyrimido[1,2- α]purin-10(3*H*)-one (M₁G) (57, 58), and the pyrimidine adduct, N⁴-(3-oxo-1-propenyl)-2'-deoxycytidine (OPdC) (59), and investigated their antigenic activity (Fig. 5A-D). We confirmed the identities of the adducts by high-resolution electro-spray ionization mass spectrometry (HR-ESI-MS) and nuclear magnetic resonance (NMR) spectroscopy (Fig. S4 and S5).

When we incubated THP-1 cells with the adducts we found that they all induced a 1.5-5 fold increase in the average expression level of MR1 on the cell, as found with other MR1-binding compounds (60) (Fig. 5A-D). THP-1 cells pulsed with a range of doses of the adducts showed that individual compounds differentially stimulated IFN- γ production by distinct MR1T-cell clones and showed cross-reactivity of some clones toward multiple compounds (Fig. 5A-D). In all cases, MR1T-cell activation by adduct-loaded THP-1 cells was fully inhibited by the addition of blocking anti-MR1 monoclonal antibodies (mAbs), confirming MR1 restriction of adduct recognition (Fig. 5A-D).

We further confirmed the stimulatory capacity of these antigens by extending the activation assays to include additional MR1T-cell clones expressing different TCRs: of the fourteen randomly selected MR1T clones tested, eight reacted to at least one ligand. M₃ADE significantly stimulated six clones, OPdA eight clones, M₁G and OPdC three clones each (Fig. 5F). Again, distinct patterns of adduct recognition resulted in a wide range of IFN- γ production levels by MR1T cells, alongside MR1T-clonal cross-reactivity. We also found that M₁G or M₃ADE when added to clones QY1A16 and QY1C3, respectively, resulted in significant decreased response to THP-1 cells, suggesting that both adducts compete with endogenous stimulatory molecules in THP-1 cells. These findings also suggested that some MR1T clones recognize antigens different from the tested ones. Of note, THP-1 cells pulsed with the synthetic compounds did not stimulate the MAIT-cell clone MRC25 (Fig. S2F).

We then used plastic-bound recombinant MR1 molecules loaded with synthetic antigens to assess whether THP-1 cell processing of the adducts was needed for MR1T-cell recognition: we found that it was not (Fig. 5E). Thus, carbonyl-nucleobase adducts directly bind MR1 without further modifications and stimulate MR1T-cells.

Next, we asked how MR1T clones that recognized distinct synthetic compounds would react to different tumor cell lines expressing physiologically low levels of surface MR1 (Fig. S3B). We performed all the experiments in the presence or absence of blocking anti-MR1 mAbs to discriminate MR1-dependent and MR1-independent reactivity. Clone AVA34, activated by M₃ADE and OPdA, specifically reacted to KMOE-2 tumor cell line only; clone QY1A16, stimulated by M₃ADE, OPdA and OPdC, selectively recognized H460 and Mel JuSo cells; while clones AC1A4, responding to M₃ADE, OPdA and M₁G, and TC5A87, activated by OPdA and OPdC, reacted in an MR1-dependent manner to all six tested tumor cell lines (Fig. 5G). Of note, the tumor lines were derived from different tissues (Fig. S3C). The tested MR1T clones did not react to leukocytes from healthy individuals, with the exception of AVA34 and TC5A87 clones that released very low, albeit significant, quantities of IFN- γ (Fig. 5G).

In conclusion, MR1T cells recognize compounds containing intact carbonyl adducts of nucleobases presented on MR1 molecules. Some MR1T clones seem to be specific for one or another adduct; others show a degree of cross-reactivity. The broad recognition and the broad distribution of nucleobase adducts in many cancers (55) may justify the broad reactivity of MR1T cells to tumors derived from different tissues.

MR1 tetramers loaded with nucleobase adduct-containing metabolites detect MR1T cells

To identify and characterize *ex vivo* MR1T cells reactive to nucleobase adduct-containing metabolites, we generated MR1 tetramers loaded with the synthesized adducts. We focused on M₃ADE, as it was the most efficient in increasing MR1 surface expression, showed the highest potency, and was recognized by 6/14 tested MR1T-cell clones. We performed several experiments to confirm proper protein refolding and MR1-M₃ADE tetramer-specific MR1T-cell staining. MR1-M₃ADE monomers stimulated clone DGB129 in plate-bound assays (Fig. S6A-B), and fluorescently-labeled tetramerized MR1-M₃ADE monomers stained the MR1T-cell clone AC1A4 (Fig. S6C), consistent with its M₃ADE reactivity (Fig. 5A); whereas the tetramers failed to bind a canonical MAIT clone (MRC25; Fig. S6C and S6F).

For further validation, the MR1-M₃ADE tetramers were used to isolate MR1T cells from peripheral blood mononuclear cells (PBMCs) and establish a novel series of clones. These clones were able to bind MR1-M₃ADE tetramers but not MR1-5-OP-RU or MR1-6-FP tetramers, demonstrated by the representative clone AVA34 (Fig. S6D). Abrogation of binding using an anti-TCRV β monoclonal antibody, further confirmed TCR specificity of the tetramer staining (Fig. S6E). In addition, the MAIT clone MRC25 was labeled with the MR1-5-OP-RU tetramer but not with the MR1-M₃ADE tetramer (Fig. S6F). Importantly, clone AVA34 reacted

310 to THP-1 cells loaded with M₃ADE, while failed to respond to other tested antigens (Fig. S6G). Thus, these experiments validated the capacity of the MR1-M₃ADE tetramer to bind M₃ADE-specific T cells.

We next asked whether MR1-M₃ADE tetramers could identify specific MR1T cells *ex vivo* in the blood of healthy donors. Screening of PBMCs from nine healthy individuals revealed
315 tetramer-positive cells in all individuals with the frequencies ranging between 0.005 and 0.077% (median = 0.01%) of total CD3⁺ cells (Fig. 6A, S6H, I).

Alongside the marked inter-individual differences in the frequency of MR1-M₃ADE tetramer-binding T cells, we also uncovered notable phenotypic variations across the tetramer-positive populations. While most cells expressed CD8 α (range 38-97%, median 75%),
320 one donor displayed a distinct CD4⁺ cell population that accounted for 57% of all tetramer-positive cells (Fig. 6B, D). In another donor, the CD4⁺/CD8⁻ population accounted for 33% of tetramer-positive cells (Fig. 6B, D). According to CD45RA and CCR7 expression, we also observed large donor-specific differences in the frequency of tetramer-positive naïve cells (range 8-55%, median 29%), central memory (range 7-48%, median 22%), effector memory
325 (range 5-81%, median 23%), and T_{EMRA} (range <1-42%, median 3.8%) (Fig. 6C, D).

Thus, MR1T cells that recognize carbonyl-nucleoside-adducts are present in the blood of healthy individuals; they display a heterogeneous phenotype, and potentially undergo phenotypic differentiation in a donor-specific manner.

330 **MR1T cells reactive to nucleobase adduct-containing metabolites infiltrate tumor tissue**

As all tested individuals possessed potentially M₃ADE-reactive T cells in their blood, we next asked whether M₃ADE-reactive T cells could be detected in tumor samples. We isolated TILs from non-small cell lung cancer biopsies from two patients and co-cultured them with A375-MR1 cells loaded with M₃ADE, in order to expand MR1T cells. After expansion, we were able
335 to detect MR1-M₃ADE tetramer-binding cells in TIL co-cultures from both patients (Fig. 6E, F). In donor #840, these cells were CD4⁺ (52.6%) or CD8⁺ (42%); whereas in donor #895 they were largely CD8⁺ (93.5%) (Fig. 6E, F). To confirm their antigen specificity, we sorted MR1-M₃ADE tetramer-positive cells and stimulated them again with either A375- β 2mKO cells, A375-MR1 or A375-MR1 loaded with M₃ADE. We measured the extent of MR1T-cell activation via TCR
340 downregulation, which occurred in the presence of A375-MR1 cells and M₃ADE-loaded A375-MR1 cells, and was prevented by anti-MR1 blocking mAbs (Fig. 6G). We further confirmed MR1-dependent activation of tetramer-positive T cells by measuring IFN- γ release. Cells from donor 840 responded to A375-MR1 cells irrespective of the presence of M₃ADE (Fig. 6H),

whereas cells from donor 895 reacted only to M₃ADE-loaded A375-MR1 cells (Fig. 6I); in both
345 cases the response was inhibited by anti-MR1 mAbs (Fig. 6H, I). These data confirm the
presence of M₃ADE-reactive T cells within the tumor environment, thus indicating a potential
role for MR1T cells in tumor immunity.

Discussion

350 In this study, we have identified nucleobase adduct-containing metabolites as self-antigens
capable of stimulating human T lymphocytes recognizing MR1-expressing tumor cells.
Previous studies showed that MR1T cells react to unique compounds fractionated from tumor
cells, suggesting distinct antigen specificity (1). Here, we confirm those data and extend them
to show that structurally-diverse nucleobase adduct-containing compounds bind MR1 and
355 stimulate individual MR1T cells. Both purines and pyrimidines form antigenic adducts and
different carbonyls participate in their generation, confirming that MR1 is a molecule with
versatile antigen-binding capacity (9).

Carbonyls accumulate as a consequence of different metabolic alterations including
glycolysis, lipid peroxidation, the metabolism of biogenic amines, vitamins and steroids, as
360 well as biotransformation of environmental agents and drugs (36, 52, 61). How many
carbonyls participate in the generation of MR1T antigens warrants further investigation: this
is an important question as the number and diversity of carbonyl species involved may
determine the size and variety of the MR1T cell antigen repertoire. It is of note that some
MR1T clones did not react to any of the antigens tested here, whereas still responded to A375-
365 MR1 cells, suggesting the MR1T cell antigen repertoire can be quite heterogeneous.

While carbonyl accumulation is important, it is not sufficient to stimulate MR1T cells, as
concomitant accumulation of purines and pyrimidines must occur within the target cells. The
structures of modified nucleobases are suited to MR1 binding and resemble those of other
MR1 ligands (62). As purines and pyrimidines are not immunogenic *per se*, the moieties
370 contributed by carbonyls are important to form antigenic complexes with MR1, likewise some
bacteria-derived MAIT-stimulatory compounds (8). Future structural studies will certainly
inform on the mode used by these antigens to bind MR1 and engage specific MR1T TCRs.

Important players in the generation of immunogenic nucleobase adducts are ROS — side
products of oxidative phosphorylation that promote tumor induction and tumor cell
375 proliferation (63). ROS also promote lipid peroxidation leading to carbonyl generation, thus
exerting indirect effects on accumulation of MR1T cell antigens. Indeed, treatment with ROS-
inducing drugs enhanced MR1T-target cell stimulatory capacity, which in turn was dampened

by the addition of ROS scavengers. Thus, MR1T-cell antigens are derived by combined alterations of multiple metabolic pathways, leading to the accumulation of nucleobases, carbonyls and ROS.

An important unanswered question in the field has been the basis of the specificity of tumor cell recognition by MR1T cells. Our data provide a plausible answer to this, as most normal cells at steady state physiologically regulate the distinct metabolic pathways contributing to nucleobase adduct generation and accumulation. By contrast, tumor cells frequently become altered in many of these pathways sustaining cell proliferation, including alterations in glucose and glutamine uptake, and high cellular demand for reduced nitrogen (64). Indeed, many tumors increase the transcription of key genes involved in the *de novo* purine synthesis (65), suggesting a key role of purinosomes (66). Furthermore, tumor cells are prone to DNA damage through the generation of nucleic acid adducts formed upon DNA oxidation and interaction with end products of lipid peroxidation (67).

Importantly, in normal cells, several mechanisms are involved in scavenging highly reactive carbonyls, through which carbonyls are oxidized to carbonic acids (68) are conjugated with glutathione (69), or are reduced to less toxic alcohols (70). In our CRISPR/Cas9 screen, we identified several genes involved in these processes that were conserved (Table S1 and Data S1, S2), suggesting that they contributed to escaping MR1T-cell recognition. The relevance of these controlling mechanisms is further supported by the negative effect of carbonyl scavengers on MR1T-stimulatory capacity of tumor cells.

MR1T-cell recognition of nucleobase adduct-containing metabolites raises the question of the physiological role of these cells. It is conceivable to attribute them a potential role in surveying cells abnormally accumulating compounds responsible for DNA alterations and therefore predisposed to dangerous genetic mutations (71). The ubiquitous expression of MR1 might be instrumental to this function of cellular metabolic integrity control. Together these properties of MR1T cells make them attractive targets for immunotherapeutic use in cancer. We envisage the possibility of using selected MR1T TCR genes to redirect the specificity of cancer patient T cells toward these novel tumor-associated metabolite antigens, and therefore equip them with tumor-targeting capacity. The detection of MR1T cells within the tumor microenvironment in two lung cancer patients is promising evidence that supports the potential value of this strategy. Another possible application is the use of nucleobase adduct-containing metabolites as components of innovative anti-tumor vaccines. Importantly, the monomorphic nature of MR1 might offer the possibility to circumvent HLA-

restriction and design T-cell-based immunotherapies applicable to the entire population of cancer patients on universal basis and independent of genetic background.

In conclusion, the immune system continues to surprise us with its capacity to detect a wide repertoire of structurally-variable antigens. T-cell recognition of nucleobase adduct-
415 containing metabolites is the most recent evidence of this enormous flexibility.

REFERENCES AND NOTES

1. M. Lepore *et al.*, Functionally diverse human T cells recognize non-microbial antigens presented by MR1. *Elife* **6**, (2017).
420
2. M. D. Crowther *et al.*, Genome-wide CRISPR-Cas9 screening reveals ubiquitous T cell cancer targeting via the monomorphic MHC class I-related protein MR1. *Nat Immunol* **21**, 178-185 (2020).
3. A. Vacchini, A. Chancellor, J. Spagnuolo, L. Mori, G. De Libero, MR1-Restricted T Cells Are Unprecedented Cancer Fighters. *Front Immunol* **11**, 751 (2020).
425
4. E. Treiner *et al.*, Selection of evolutionarily conserved mucosal-associated invariant T cells by MR1. *Nature* **422**, 164-169 (2003).
5. L. Le Bourhis *et al.*, Antimicrobial activity of mucosal-associated invariant T cells. *Nature Immunology* **11**, 701-708 (2010).
6. M. C. Gold *et al.*, Human Mucosal Associated Invariant T Cells Detect Bacterially Infected Cells. *PLoS Biology* **8**, e1000407 (2010).
430
7. L. Kjer-Nielsen *et al.*, MR1 presents microbial vitamin B metabolites to MAIT cells. *Nature* **491**, 717-723 (2012).
8. A. J. Corbett *et al.*, T-cell activation by transitory neo-antigens derived from distinct microbial pathways. *Nature* **509**, 361-365 (2014).
435
9. A. N. Keller *et al.*, Drugs and drug-like molecules can modulate the function of mucosal-associated invariant T cells. *Nat Immunol* **18**, 402-411 (2017).
10. M. J. Harriff *et al.*, MR1 displays the microbial metabolome driving selective MR1-restricted T cell receptor usage. *Science Immunology* **3**, eaao2556 (2018).
11. M. Salio *et al.*, Ligand-dependent downregulation of MR1 cell surface expression. *Proc Natl Acad Sci U S A* **117**, 10465-10475 (2020).
440
12. N. A. Gherardin *et al.*, Diversity of T Cells Restricted by the MHC Class I-Related Molecule MR1 Facilitates Differential Antigen Recognition. *Immunity* **44**, 32-45 (2016).
13. D. Hanahan, R. A. Weinberg, Hallmarks of cancer: the next generation. *Cell* **144**, 646-674 (2011).
445
14. M. G. Vander Heiden, L. C. Cantley, C. B. Thompson, Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation. *Science* **324**, 1029-1033 (2009).
15. O. Shalem *et al.*, Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84-87 (2014).
450
16. T. Hart *et al.*, High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. *Cell* **163**, 1515-1526 (2015).

- 455 17. M. F. Pech *et al.*, Systematic identification of cancer cell vulnerabilities to natural killer cell-mediated immune surveillance. *Elife* **8**, (2019).
18. M. Ashburner *et al.*, Gene Ontology: tool for the unification of biology. *Nature Genetics* **25**, 25 (2000).
19. The Gene Ontology Consortium, The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Research* **47**, D330-D338 (2018).
- 460 20. E. Brunk *et al.*, Recon3D enables a three-dimensional view of gene variation in human metabolism. *Nature Biotechnology* **36**, 272-281 (2018).
21. M. Uhr, J. Stelling, Structural Sensitivity Analysis of Metabolic Networks. *IFAC Proceedings Volumes* **41**, 15879-15884 (2008).
- 465 22. O. Oeyas *et al.*, Model-based integration of genomics and metabolomics reveals SNP functionality in Mycobacterium tuberculosis. *bioRxiv*, 555763 (2019).
23. H. C. Losey, A. J. Ruthenburg, G. L. Verdine, Crystal structure of Staphylococcus aureus tRNA adenosine deaminase TadA in complex with RNA. *Nat Struct Mol Biol* **13**, 153-159 (2006).
- 470 24. H. Sun *et al.*, Molecular cloning and characterization of a novel muscle adenylosuccinate synthetase, AdSSL1, from human bone marrow stromal cells. *Mol Cell Biochem* **269**, 85-94 (2005).
25. M. Z. Cader *et al.*, FAMIN Is a Multifunctional Purine Enzyme Enabling the Purine Nucleotide Cycle. *Cell* **180**, 815 (2020).
- 475 26. J. P. Richard, Kinetic parameters for the elimination reaction catalyzed by triosephosphate isomerase and an estimation of the reaction's physiological significance. *Biochemistry* **30**, 4581-4585 (1991).
27. B. Stratmann, B. Goldstein, P. J. Thornalley, N. Rabbani, D. Tschoepe, Intracellular Accumulation of Methylglyoxal by Glyoxalase 1 Knock Down Alters Collagen Homeostasis in L6 Myoblasts. *Int J Mol Sci* **18**, (2017).
- 480 28. M. Frischmann, C. Bidmon, J. Angerer, M. Pischetsrieder, Identification of DNA Adducts of Methylglyoxal. *Chemical Research in Toxicology* **18**, 1586-1592 (2005).
29. M. J. Nokin *et al.*, Methylglyoxal, a glycolysis side-product, induces Hsp90 glycation and YAP-mediated tumor growth and metastasis. *Elife* **5**, (2016).
- 485 30. P. J. Thornalley *et al.*, Antitumour activity of S-p-bromobenzylglutathione cyclopentyl diester in vitro and in vivo. Inhibition of glyoxalase I and induction of apoptosis. *Biochem Pharmacol* **51**, 1365-1372 (1996).
- 490 31. A. Al-Timari, K. T. Douglas, Inhibition of mammalian glyoxalase I (lactoylglutathione lyase) by N-acylated S-blocked glutathione derivatives as a probe for the role of the N-site of glutathione in glyoxalase I mechanism. *Biochim Biophys Acta* **870**, 160-168 (1986).

32. T. J. Franklin, J. M. Cook, The inhibition of nucleic acid synthesis by mycophenolic acid. *Biochem J* **113**, 515-524 (1969).
33. T. Podzuweit, P. Nennstiel, A. Muller, Isozyme selective inhibition of cGMP-stimulated cyclic nucleotide phosphodiesterases by erythro-9-(2-hydroxy-3-nonyl) adenine. *Cell Signal* **7**, 733-738 (1995).
495
34. R. Quintens *et al.*, Mice deficient in the respiratory chain gene *Cox6a2* are protected against high-fat diet-induced obesity and insulin resistance. *PLoS One* **8**, e56719 (2013).
35. S. Miwa *et al.*, Low abundance of the matrix arm of complex I in mitochondria predicts longevity in mice. *Nat Commun* **5**, 3837 (2014).
500
36. R. J. Mailloux, S. L. McBride, M. E. Harper, Unearthing the secrets of mitochondrial ROS and glutathione in bioenergetics. *Trends Biochem Sci* **38**, 592-602 (2013).
37. J. D. Hayes, J. U. Flanagan, I. R. Jowsey, Glutathione transferases. *Annu Rev Pharmacol Toxicol* **45**, 51-88 (2005).
38. N. Allocati, M. Masulli, C. Di Ilio, L. Federici, Glutathione transferases: substrates, inhibitors and pro-drugs in cancer and neurodegenerative diseases. *Oncogenesis* **7**, 8 (2018).
505
39. M. R. Engle *et al.*, Physiological role of mGSTA4-4, a glutathione S-transferase metabolizing 4-hydroxynonenal: generation and analysis of mGsta4 null mouse. *Toxicol Appl Pharmacol* **194**, 296-308 (2004).
510
40. J. Alexandre *et al.*, Accumulation of hydrogen peroxide is an early and crucial step for paclitaxel-induced cancer cell death both in vitro and in vivo. *Int J Cancer* **119**, 41-48 (2006).
41. S. Luanpitpong *et al.*, Mitochondrial superoxide mediates doxorubicin-induced keratinocyte apoptosis through oxidative modification of ERK and Bcl-2 ubiquitination. *Biochem Pharmacol* **83**, 1643-1654 (2012).
515
42. M. N. Benckroun, P. Pourquier, B. Schott, J. Robert, Doxorubicin-induced lipid peroxidation and glutathione peroxidase activity in tumor cell lines selected for resistance to doxorubicin. *Eur J Biochem* **211**, 141-146 (1993).
43. S. Heumuller *et al.*, Apocynin is not an inhibitor of vascular NADPH oxidases but an antioxidant. *Hypertension* **51**, 211-217 (2008).
520
44. O. W. Griffith, A. Meister, Glutathione: interorgan translocation, turnover, and metabolism. *Proc Natl Acad Sci U S A* **76**, 5606-5610 (1979).
45. C. Lennicke, J. Rahn, R. Lichtenfels, L. A. Wessjohann, B. Seliger, Hydrogen peroxide - production, fate and role in redox signaling of tumor cells. *Cell Commun Signal* **13**, 39 (2015).
525
46. A. Ayala, M. F. Munoz, S. Arguelles, Lipid peroxidation: production, metabolism, and signaling mechanisms of malondialdehyde and 4-hydroxy-2-nonenal. *Oxid Med Cell Longev* **2014**, 360438 (2014).

- 530 47. P. J. O'Brien, A. G. Siraki, N. Shangari, Aldehyde sources, metabolism, molecular toxicity mechanisms, and possible effects on human health. *Crit Rev Toxicol* **35**, 609-662 (2005).
48. A. K. Chaudhary *et al.*, Detection of endogenous malondialdehyde-deoxyguanosine adducts in human liver. *Science* **265**, 1580-1582 (1994).
- 535 49. A. Munnia, M. E. Amasio, M. Peluso, Exocyclic malondialdehyde and aromatic DNA adducts in larynx tissues. *Free Radic Biol Med* **37**, 850-858 (2004).
50. T. Matsuda *et al.*, Lipid peroxidation-induced DNA adducts in human gastric mucosa. *Carcinogenesis* **34**, 121-127 (2013).
- 540 51. M. Peluso *et al.*, Breast fine-needle aspiration malondialdehyde deoxyguanosine adduct in breast cancer. *Free Radic Res* **45**, 477-482 (2011).
52. F. Gentile *et al.*, DNA damage by lipid peroxidation products: implications in cancer, inflammation and autoimmunity. *AIMS Genet* **4**, 103-137 (2017).
53. R. Brigelius-Flohe, L. Flohe, Regulatory Phenomena in the Glutathione Peroxidase Superfamily. *Antioxid Redox Signal* **33**, 498-516 (2020).
- 545 54. M. Colzani *et al.*, Reactivity, Selectivity, and Reaction Mechanisms of Aminoguanidine, Hydralazine, Pyridoxamine, and Carnosine as Sequestering Agents of Reactive Carbonyl Species: A Comparative Study. *ChemMedChem* **11**, 1778-1789 (2016).
55. G. P. Voulgaridou, I. Anestopoulos, R. Franco, M. I. Panayiotidis, A. Pappa, DNA damage induced by endogenous aldehydes: current state of knowledge. *Mutat Res* **711**, 13-27 (2011).
- 550 56. K. Stone, M. B. Ksebati, L. J. Marnett, Investigation of the adducts formed by reaction of malondialdehyde with adenosine. *Chemical Research in Toxicology* **3**, 33-38 (1990).
57. L. J. Marnett, Lipid peroxidation—DNA damage by malondialdehyde. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **424**, 83-95 (1999).
- 555 58. S. A. Yates, N. M. Dempster, M. F. Murphy, S. A. Moore, Quantitative analysis of malondialdehyde-guanine adducts in genomic DNA samples by liquid chromatography/tandem mass spectrometry: LC/MS/MS analysis of malondialdehyde-guanine adducts. *Rapid Communications in Mass Spectrometry* **31**, 762-770 (2017).
- 560 59. K. Stone, A. Uzieblo, L. J. Marnett, Studies of the reaction of malondialdehyde with cytosine nucleosides. *Chemical Research in Toxicology* **3**, 467-472 (1990).
60. H. E. G. McWilliam *et al.*, The intracellular pathway for the presentation of vitamin B-related antigens by the antigen-presenting molecule MR1. *Nature Immunology* **17**, 531-537 (2016).
- 565 61. R. De Bont, N. van Larebeke, Endogenous DNA damage in humans: a review of quantitative data. *Mutagenesis* **19**, 169-185 (2004).

- 570 62. A. J. Corbett, W. Awad, H. Wang, Z. Chen, Antigen Recognition by MR1-Reactive T Cells; MAIT Cells, Metabolites, and Remaining Mysteries. *Front Immunol* **11**, 1961 (2020).
63. P. T. Schumacker, Reactive oxygen species in cancer cells: live by the sword, die by the sword. *Cancer Cell* **10**, 175-176 (2006).
64. Natalya N. Pavlova, Craig B. Thompson, The Emerging Hallmarks of Cancer Metabolism. *Cell Metabolism* **23**, 27-47 (2016).
- 575 65. R. Nilsson *et al.*, Metabolic enzyme expression highlights a key role for MTHFD2 and the mitochondrial folate pathway in cancer. *Nat Commun* **5**, 3128 (2014).
66. A. M. Pedley, S. J. Benkovic, A New View into the Regulation of Purine Metabolism: The Purinosome. *Trends Biochem Sci* **42**, 141-154 (2017).
- 580 67. D. Pluskota-Karwatka, Modifications of nucleosides by endogenous mutagens-DNA adducts arising from cellular processes. *Bioorg Chem* **36**, 198-213 (2008).
68. V. Vasiliou, A. Pappa, D. R. Petersen, Role of aldehyde dehydrogenases in endogenous and xenobiotic metabolism. *Chem Biol Interact* **129**, 1-19 (2000).
69. J. D. Hayes, L. I. McLellan, Glutathione and glutathione-dependent enzymes represent a co-ordinately regulated defence against oxidative stress. *Free Radic Res* **31**, 273-300 (1999).
- 585 70. L. Zhong *et al.*, Aldo-keto reductase family 1 B10 protein detoxifies dietary and lipid-derived alpha, beta-unsaturated carbonyls at physiological levels. *Biochem Biophys Res Commun* **387**, 245-250 (2009).
71. L. J. Marnett, J. P. Plataras, Endogenous DNA damage and mutation. *Trends Genet* **17**, 214-221 (2001).
- 590 72. M. Lepore *et al.*, Parallel T-cell cloning and deep sequencing of human MAIT cells reveal stable oligoclonal TCRbeta repertoire. *Nat Commun* **5**, 3866 (2014).
73. J. Joung *et al.*, Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. *Nat Protoc* **12**, 828-863 (2017).
- 595 74. M. Schmalzer *et al.*, Modulation of bacterial metabolism by the microenvironment controls MAIT cell stimulation. *Mucosal Immunology*, (2018).
75. H. Seto, T. Takesue, T. Ikemura, Reaction of Malonaldehyde with Nucleic Acid. II. Formation of Fluorescent Pyrimido[1,2-a]purin-10(3H)-one Mononucleotide. *Bulletin of the Chemical Society of Japan* **58**, 3431-3435 (1985).
- 600 76. M. Hadley, H. H. Draper, Isolation of a guanine-malondialdehyde adduct from rat and human urine. *Lipids* **25**, 82 (1990).
77. J. Szekely *et al.*, "One-Pot" Syntheses of Malondialdehyde Adducts of Nucleosides. *Nucleosides, Nucleotides and Nucleic Acids* **27**, 103-109 (2008).

- 605 78. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120 (2014).
79. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359 (2012).
80. R Development Core Team. (R Foundation for Statistical Computing, Vienna, Austria, 2018).
- 610 81. Z. Dai *et al.*, edgeR: a versatile tool for the analysis of shRNA-seq and CRISPR-Cas9 genetic screens. *F1000 Res* **3**, 95 (2014).
82. K. R. Sanson *et al.*, Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. *Nat Commun* **9**, 5416 (2018).
83. M. Carlson. (R package version 3.8.2, 2019).
- 615 84. S. Durinck, P. T. Spellman, E. Birney, W. Huber, Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc* **4**, 1184-1191 (2009).
85. O. Oyas *et al.*, Model-based integration of genomics and metabolomics reveals SNP functionality in *Mycobacterium tuberculosis*. *Proc Natl Acad Sci U S A* **117**, 8494-8502 (2020).
- 620 86. J. D. Orth, I. Thiele, B. O. Palsson, What is flux balance analysis? *Nat Biotechnol* **28**, 245-248 (2010).
87. J. Schellenberger, N. E. Lewis, B. O. Palsson, Elimination of thermodynamically infeasible loops in steady-state metabolic models. *Biophys J* **100**, 544-553 (2011).
- 625 88. S. H. J. Chan, L. Wang, S. Dash, C. D. Maranas, Accelerating flux balance calculations in genome-scale metabolic models by localizing the application of loopless constraints. *Bioinformatics* **34**, 4248-4255 (2018).

630 **Acknowledgements:**

We thank Verena Schaefer, Petra Herzig and Florine Winter for contributing to experiments, Feng Zhang for Addgene Human GeCKO v2 knockout library and plasmid 52962. The MR1 tetramer technology was developed jointly by James McCluskey, Jamie Rossjohn and David Fairlie, and 5-OP-RU- and 6-FP-hMR1 tetramers were produced by the NIH Tetramer Core Facility as permitted to be distributed by the University of Melbourne. We wish to thank Lucy Robinson and Jessica Tamanini of Insight Editing London for assistance in preparing the manuscript. **Funding:** Work in De Libero's group was supported by grants of Swiss National Foundation (310030-173240 and 310030B-192828), Swiss Cancer League (KFS-4707-02-2019), Cancer League beider Basel (KLbB-4779-02-2019), D-BSSE ETH Zürich (PMB-02-17) and by the University and University Hospital of Basel. **Author Contributions** Conceptualization and Methodology, L.M. and G.D.L.; Investigation, A.V., A.C., J.S. Q.Y., D.J., A.B., C.D.G., M.P. and M.L.; Formal analysis, O.Ø., J.St.; Resources, A.Z., G.D.L.; Data curation, A.V., A.C., J.S., D.H.; Writing original draft, A.V., A.C., J.S., G.D.L.; Writing-Review and Editing, M.L., L.M. and G.D.L.; Visualization, A.V., A.C., Q.Y., J.S., D.J.; Supervision, L.M. and G.D.L.; Funding Acquisition, G.D.L. All authors approved the manuscript. **Competing Interests:** A.V., A.C., Q.Y., J.S., M.L., L.M., G.D.L. are listed as inventors in patent applications related to MR1T-based cancer treatment. L.M. is an employee and shareholder of Matterhorn Biosciences. All other authors declare no competing interests.

650 **Supplementary Materials:**

Materials and Methods
Fig. S1 to S8
Tables S1-S2
External Data S1-S2
655 References (72-88)

Figures

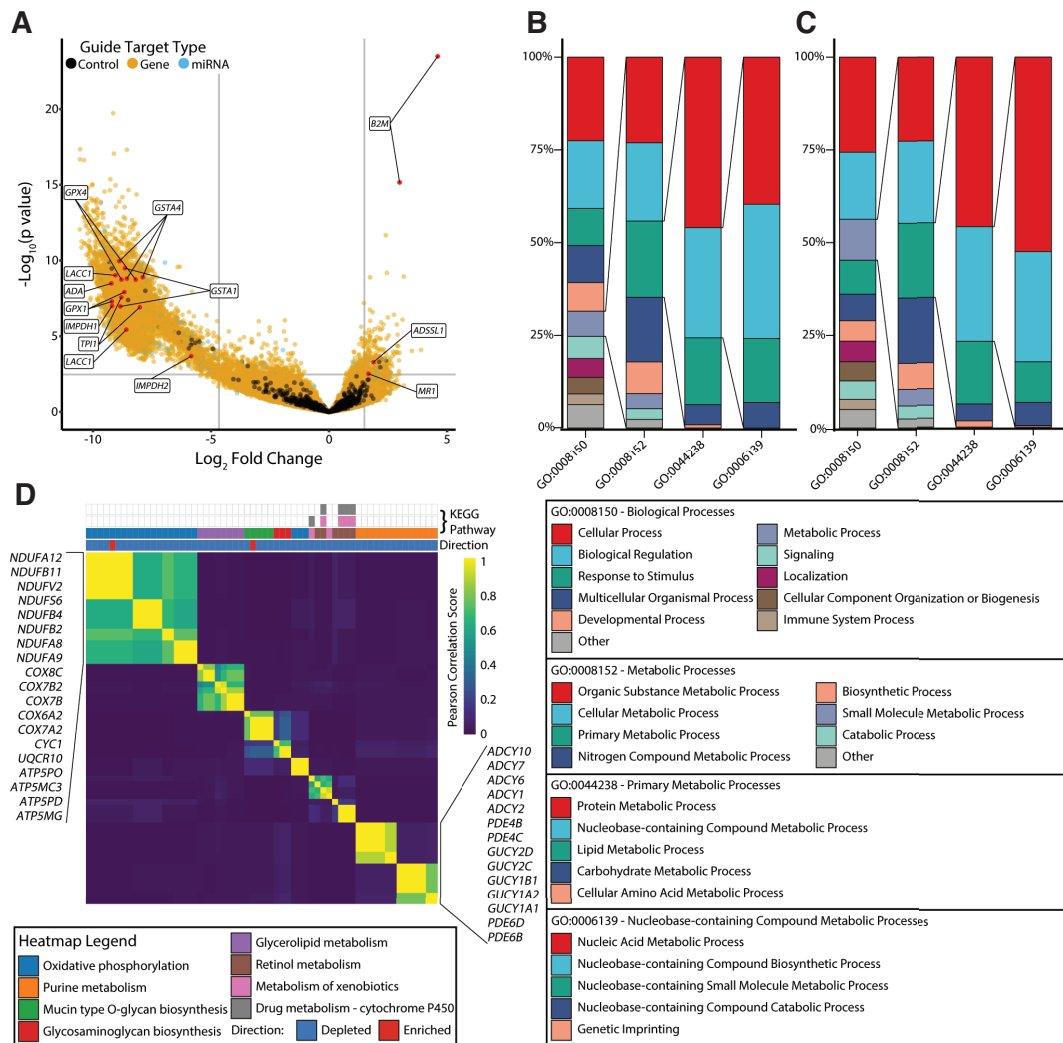


Fig. 1. Identification and analysis of gene-targets influencing the efficiency of T-cell mediated killing

660 (A) Significantly enriched and depleted guides were identified as having fold-change greater or less than the top 1% of the negative control guides (black points), respectively (thresholds indicated by vertical grey lines) and $FDR < 0.05$, here annotated as the highest p -value at which FDR is still < 0.05 (horizontal grey line). (B and C) Significantly enriched (Binomial enrichment $FDR < 0.05$) biological process GO-terms in enriched (B) and depleted (C) gene-targets. Each successive bar shows the significant child GO-terms. (D) Similarity of single-gene knockouts in the Recon3D metabolic perturbation model. See also Tables S1, Data S1 and S2.

665

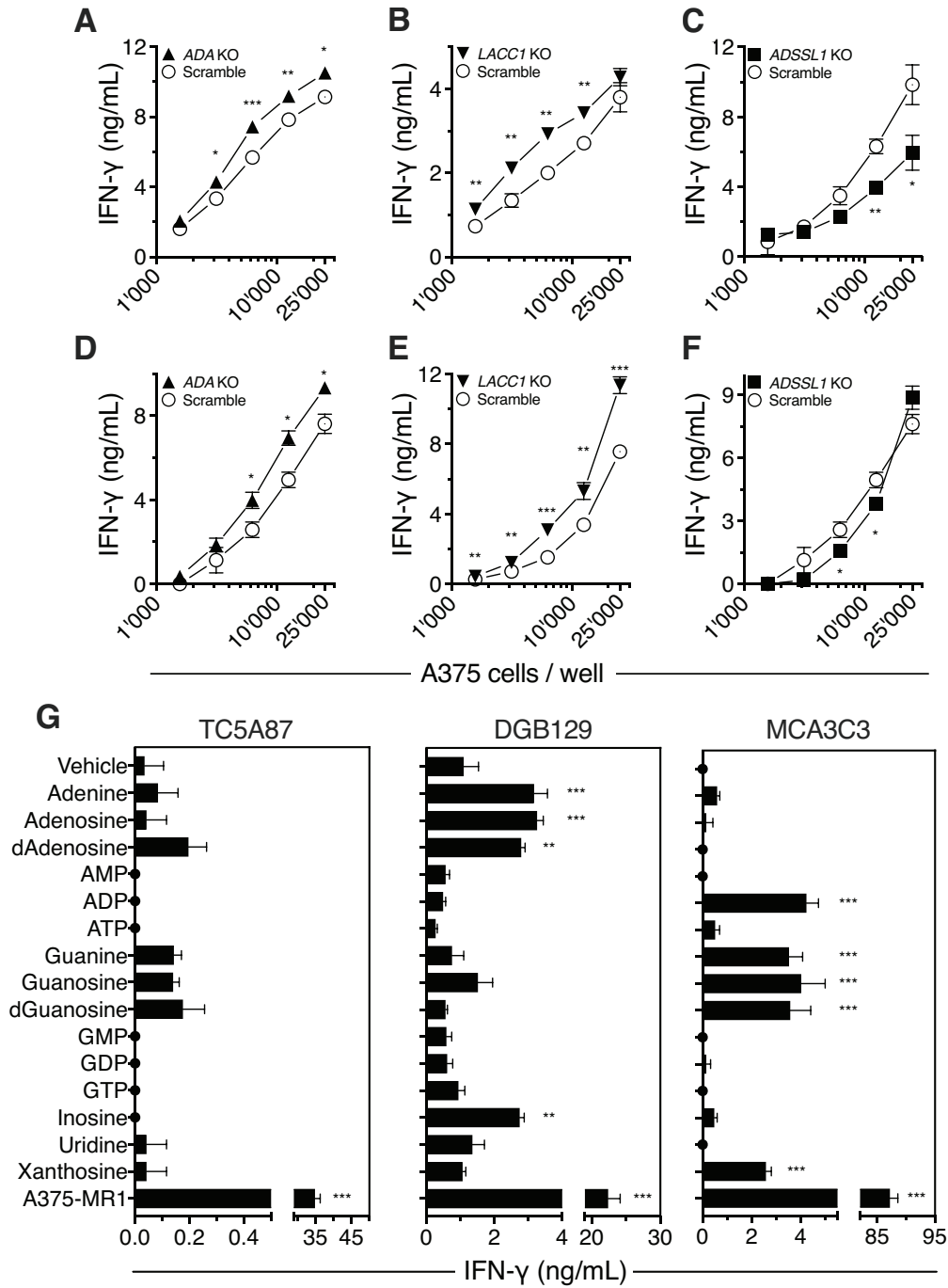


Fig. 2. Purine metabolism is involved in MR1T antigen accumulation

(A-F) MR1T clones TC5A87 (A-C) and DGB129 (D-F) reactivity against A375-MR1 cells
 670 transduced with sgRNAs targeting ADA (▲, A and D), LACC1 (▼, B and E), ADSSL1 (■, C and
 F) or scrambled sgRNA control (O, A-F).

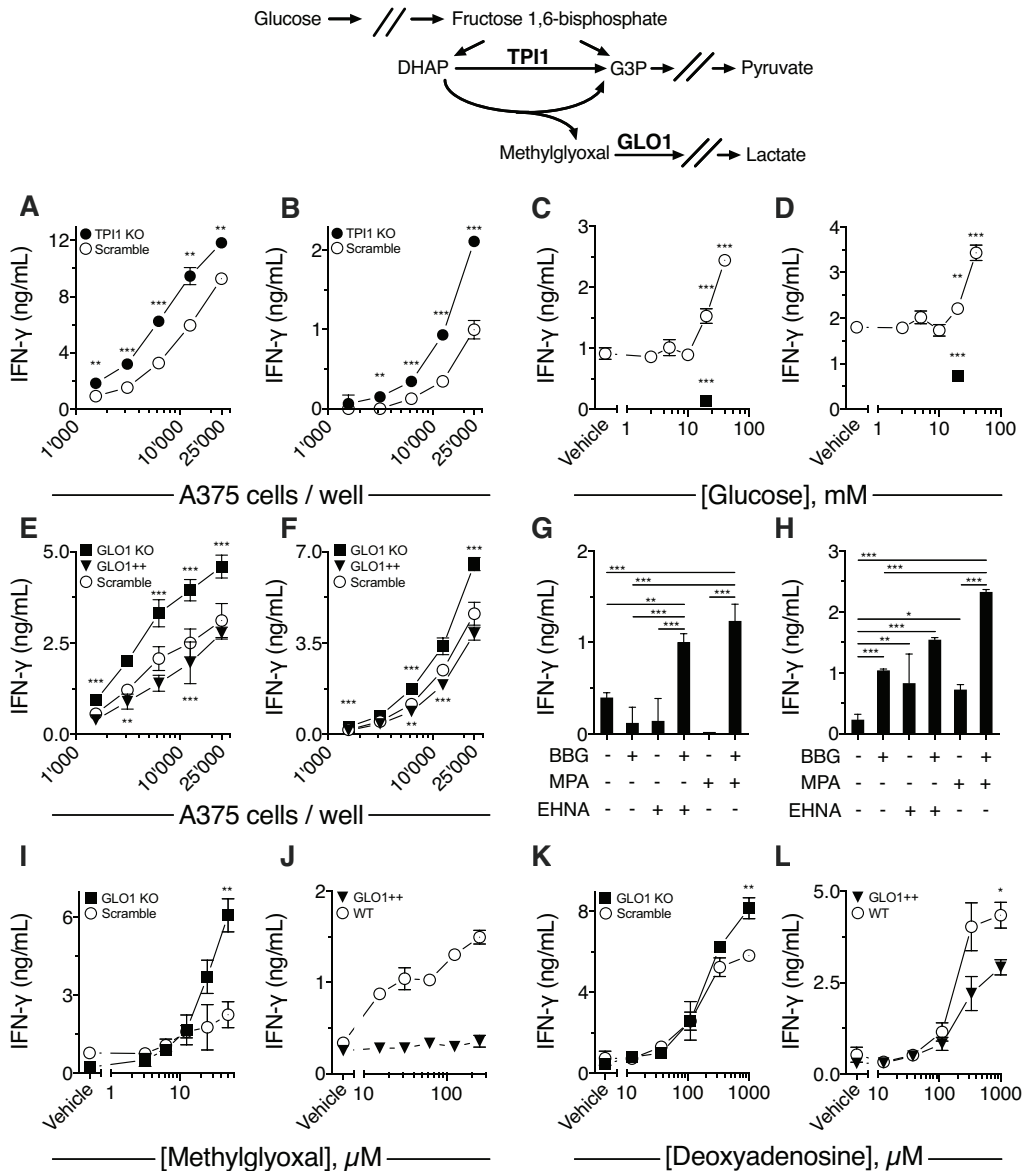
(G) Activation of MR1T clones TC5A87 (left), DGB129 (middle) and MCA3C3 (right) by THP-1
 cells pre-incubated with 250 μM of the indicated molecules or A375-MR1 or vehicle.

IFN- γ released is the mean \pm SD of triplicate cultures.

675 One representative experiment of at least three independent replicates is shown in each panel.

* $p < 0.05$, ** $p \leq 0.01$ and *** $p \leq 0.001$ compared to matching control (A-F, Multiple t-test) or compared to vehicle (G, One-way Anova with Dunnett's multiple comparison).

See also Fig. S1, S2 and S7.



680

Fig. 3. Glycolysis and methylglyoxal lead to MR1T antigen accumulation

Schematic representation of methylglyoxal generation. Dihydroxyacetone phosphate (DHAP), Glyceraldehyde 3-phosphate (G3P).

685 (A and B) Stimulation of MR1T cell clone TC5A87 (A) and DGB129 (B) with A375-MR1 cells transduced with sgRNAs targeting TPI1 (●) or scrambled control (○).

(C and D) Stimulation of MR1T cell clone TC5A87 (C) and DGB129 (D) in response to fixed A375-MR1 cells incubated with different concentrations of D-(+)-Glucose (○) or 2-deoxy-D-Glucose (■) before fixation.

26

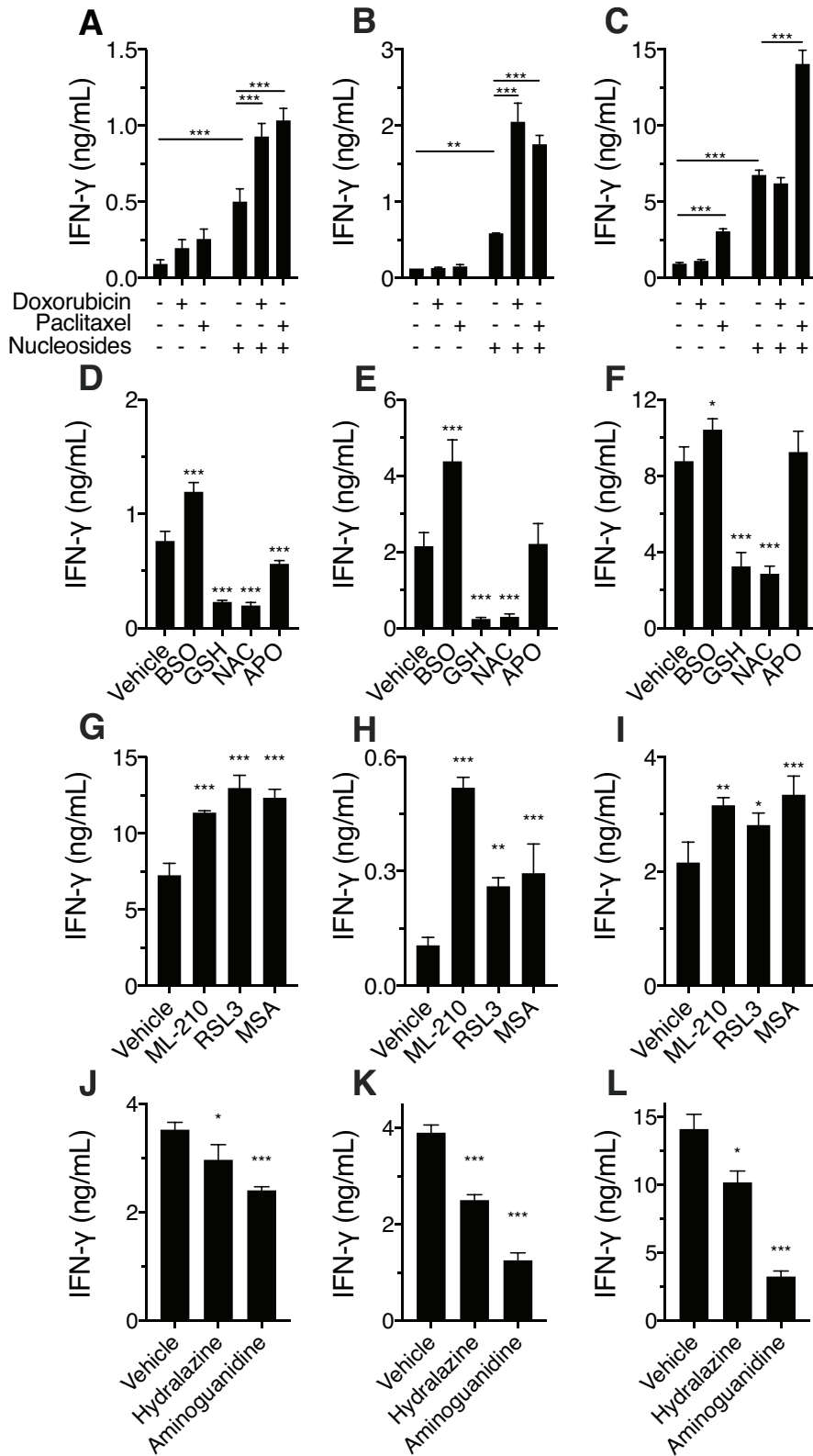
690 (E and F) Stimulation of MR1T cell clone TC5A87 (E) and DGB129 (F) with A375-MR1 cells transduced with sgRNAs against GLO1 (■), scrambled sgRNAs control (○) or a vector to overexpress GLO1 (GLO1⁺⁺ ▼).

(G and H) Stimulation of MR1T cell clone TC5A87 (G) and DGB129 (H) with THP-1 cells pre-treated with erythro-9-(2-Hydroxy-3-nonyl) adenine hydrochloride (EHNA), Mycophenolic acid (MPA) and S-Bromobenzylglutathione (BBG), alone or in combination.

695 (I-L) MR1T clone DGB129 activation in response to THP-1 cells (○), GLO1⁺⁺ (▼) and GLO1 *ko* (■) THP-1 cells, in the presence of Methylglyoxal (I, J) or deoxyadenosine (K, L).

IFN- γ released is the mean \pm SD of triplicate cultures. The data shown are representative of at least three independent experiments.

700 (A-L) * $p < 0.05$ ** $p \leq 0.01$ and *** $p \leq 0.001$. (A, B and I-L) Multiple t-test, (C and D) one-way Anova with Dunnett's multiple comparison, (E and F) two-way Anova with Dunnett's multiple comparison, (G and H) one-way Anova with Tukey's multiple comparison. See also related Fig. S1 and S2.



705

Fig. 4. Aldehyde contributing and scavenging MR1T antigen accumulation

(A-C) Stimulation of MR1T cell clone TC5A87 (A), DGB129 (B) and MCA3C3 (C) with THP-1 cells pre-treated with Doxorubicin or Paclitaxel in the absence or presence of nucleosides (dAdenosine and Guanosine).
710

(D-F) Stimulation of MR1T cell clone TC5A87 (D), MCA2B1 (E) MCA3C3 (F) with fixed A375-MR1 cells treated with Buthionine Sulfoximine (BSO), Glutathione (GSH), N-Acetylcysteine (NAC) and Apocynin (APO). (G-I) Stimulation of MR1T cell clone TC5A87 (G), DGB129 (H) MCA2B1 (I) with fixed A375-MR1 cells treated with ML-210, RSL3 and Mercaptosuccinic acid (MSA). (J-L) Stimulation of MR1T cell clone TC5A87 (J), DGB129 (K) and MCA2B1 (L) with fixed A375-MR1 cells treated with Hydralazine or Aminoguanidine. IFN- γ release is the mean \pm SD of triplicate cultures. The data shown are representative of at least three independent experiments. (A-O) * $p < 0.05$, ** $p \leq 0.01$ and *** $p \leq 0.001$. (A-C) two-way Anova with Tukey's multiple comparison, (D-L) one-way Anova with Dunnett's multiple comparison.
715

720 See also Fig. S2, S3 and S8.

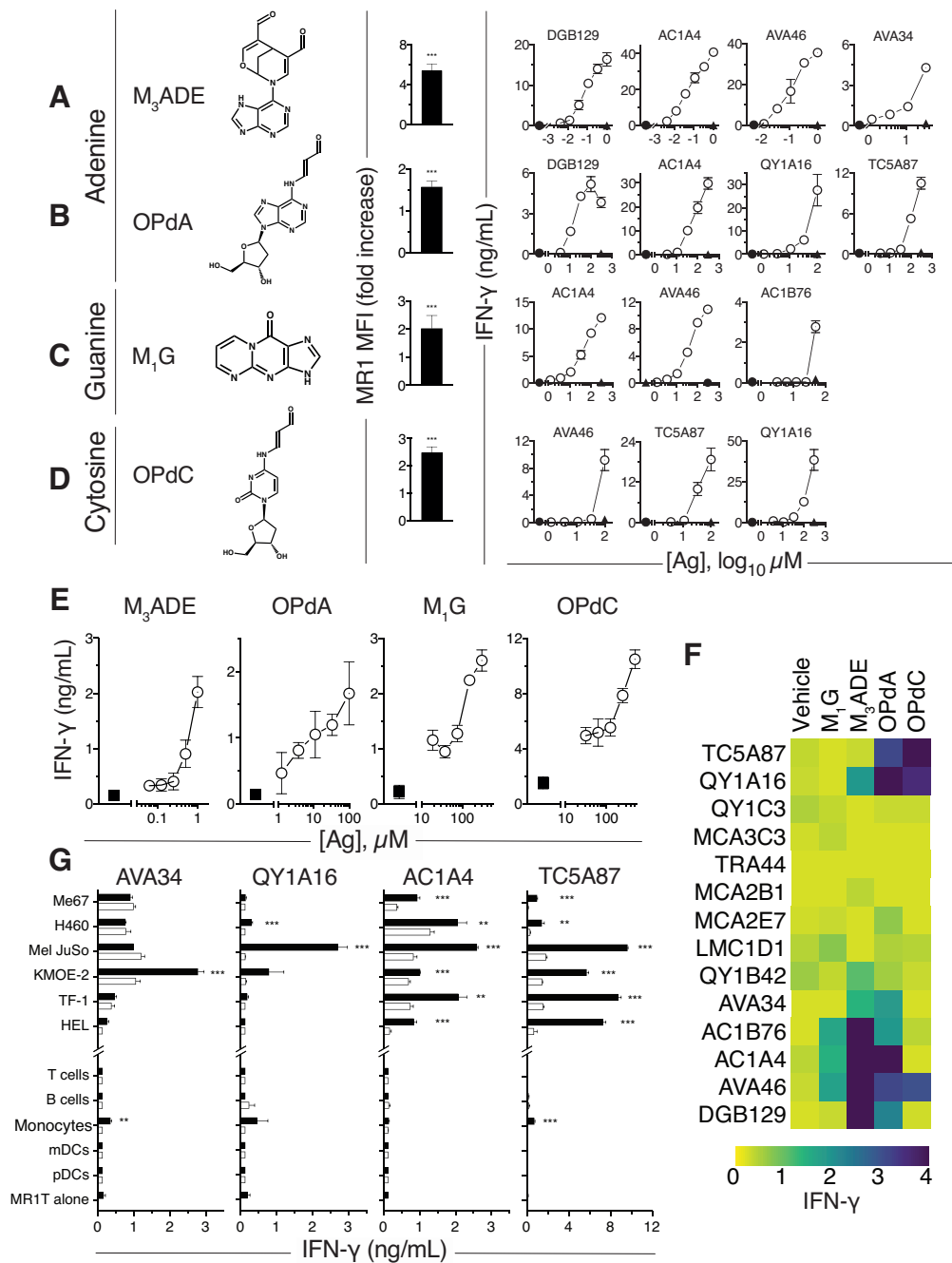


Fig. 5. Synthetic MDA nucleoside adducts stimulate MR1T cells

725 (A-D) (Left) Structure of four synthetic adducts M_3 ADE (A), OPdA (B), M_1 G (C), and OPdC (D). (Middle) Upregulation of MR1 surface expression on THP-1 cells following 6 h incubation with the indicated adducts. Fold increase \pm SD compared to cells treated with vehicle is shown. (Right) IFN- γ release response of indicated MR1T cell clones co-cultured with THP-1 cells in

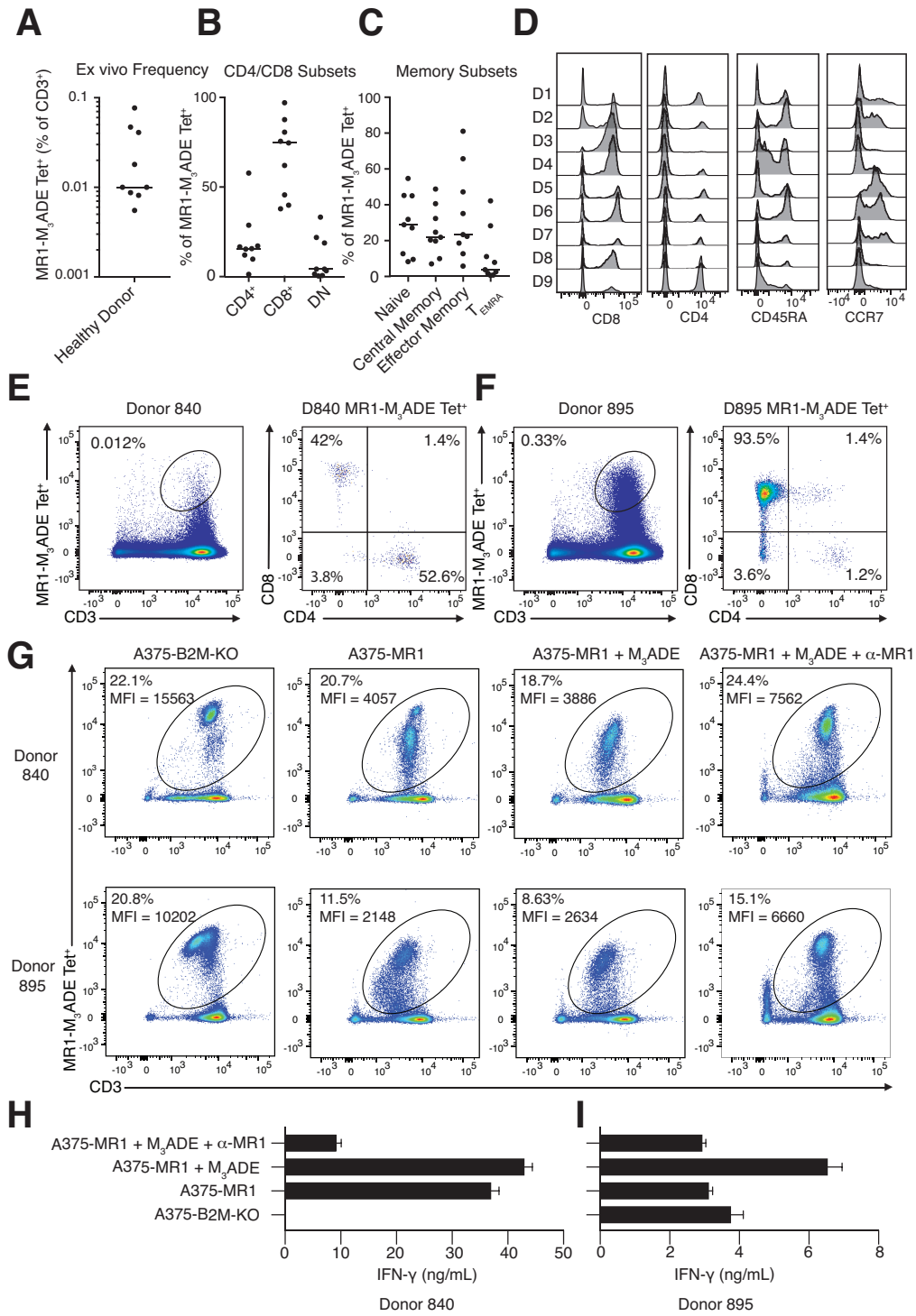
the presence (○) or absence of adducts (●). Blocking of T-cell reactivity is shown for the
730 highest antigen (Ag) dose using anti-MR1 mAbs (▲).

(E) Response of MR1T clones DGB129, QY1A16, AC1A4 and TC5A87, stimulated by plate-bound soluble MR1 loaded (○) respectively with M₃ADE, OPdA, M₁G and OPdC or not (■).

(F) Stimulation of 14 MR1T cell clones in the presence of THP-1 cells treated with M₃ADE, OPdA, M₁G, OPdC or vehicle. Heat-map reports the square root of mean IFN-γ concentration.
735 Antigen response was compared to vehicle and evaluated by multiple t-test ($p < 0.05$).

(G) MR1T cell reactivity to tumor cell lines and normal cells. Black bars depict co-cultures with indicated tumor cell line in the presence of anti-HLA-A,B,C and anti-HLA-DR mAbs. White bars represent co-cultures with both anti-HLA and anti-MR1 mAbs. n.d.=not determined

The data shown are representative of at least three independent experiments. MR1T cell
740 reactivity towards normal cells obtained from pools of healthy donors is shown. (A-D) Mean \pm SD, $n \geq 3$. (E-G) Mean \pm SD, $n \geq 2$. (G) ** $p \leq 0.01$ and *** $p \leq 0.001$, Multiple t-test. See also related Fig. S2, S3, S4 and S5.



745

Fig. 6. M₃ADE tetramer⁺T cells are heterogeneous among circulating lymphocytes and infiltrate human non-small cell lung carcinoma

(A) Summary of the *ex vivo* frequency of MR1-M₃ADE tetramer⁺ T cells from the peripheral
750 blood of nine healthy donors. Horizontal bar represents median value.

(B) Frequency of CD4⁺/CD8⁻, CD4⁻/CD8⁺ and CD4⁻/CD8⁻ (DN) T cells within MR1-M₃ADE
tetramer⁺ T cells. Horizontal bars represent median values.

(C) Percentages of indicated subsets within the MR1-M₃ADE tetramer⁺ T cell fraction.
Horizontal bars represent median values.

755 (D) Histograms of surface expression of indicated markers on T cells pre-gated as CD3⁺,
M₃ADE-MR1 tetramer⁺ from each donor (D1-D9). See also Fig. S6.

(E, F) MR1-M₃ADE tetramer⁺, CD3⁺ T cells within TILs from two patients (Donors 840 and 895)
after *in vitro* expansion. The percentage of MR1-M₃ADE tetramer⁺ T cells within total
CD3⁺ cells is displayed (Left). Dot plots show the proportion of CD4⁺/CD8⁻, CD4⁻/CD8⁺ and CD4⁻
760 /CD8⁻ T cells within MR1-M₃ADE tetramer⁺ T cells. Numbers in quadrants indicate the
percentage of each population calculated on total CD3⁺, M₃ADE-MR1 tetramer⁺ cells (Right).

(G) Dot plots show the FACS-sorted M₃ADE-tetramer enriched T-cell lines derived from TILs of
each donor activated by A375-B2M KO cells or A375-MR1, treated with M₃ADE ± blocking with
anti-MR1 mAbs. T-cell activation was measured by TCR downregulation, shown as percentage
765 and MFI of tetramer⁺ cells indicated in each plot.

(H, I) Activation of TIL lines in each of the conditions corresponding to G. IFN-γ release (mean
± SD) was measured in duplicate for each condition.

The data shown are representative of at least two independent experiments.

770

Supplementary Materials for

775 **T cell activation by nucleobase adduct-containing metabolites**

Alessandro Vacchini^{1,†}, Andrew Chancellor^{1,†}, Qinmei Yang^{1,†}, Julian Spagnuolo^{1,†}, Daniel Joss², Ove Øyås^{3,4}, Aisha Beshirova¹, Corinne De Gregorio¹, Michael Pfeffer², Alfred Zippelius⁵, Jörg Stelling³, Daniel Häussinger², Marco Lepore^{1,6}, Lucia Mori¹, Gennaro De
780 Libero^{1*}

*Corresponding author: gennaro.delibero@unibas.ch

785 **This PDF file includes:**

Materials and Methods

Figs. S1 to S8

Tables S1 to S2

790 Captions for Data S1 to S2

Other Supplementary Materials for this manuscript include the following:

Data S1 to S2 (Excel files)

- 795
- Data S1, Significantly enriched sgRNAs classification according to KEGG pathway, Related to Figure 1 and Table S1.
 - Data S2, Significantly depleted sgRNAs classification according to KEGG pathway, Related to Figure 1 and Table S1.

800

Materials and Methods

805 Human blood samples

Blood and tissue specimens for T-cell cloning, FACS analysis and antigen-presentation assays were obtained from the University Hospital Basel. The study was approved by the local ethical review board (EKNZ, Ethics Committee North-West & Central Switzerland, EKNZ 2017-01888 and EK128/13), and all patients and healthy donors consented in writing to the analysis of
810 their samples.

Cell Culture

A375-MR1 cells were generated as previously described (1). The H460 (ATCC #HTB-177), HEK-293T (ATCC #CRL-3216), HEL (ATCC #CVCL_0001), KMOE-2 (ATCC #CVCL_1332), Me67 (ATCC
815 #CVCL_VT20), Me1 JUSO (ATCC #CVCL_1403), TF-1 (ATCC #CRL-2003) and THP-1 (ATCC #TIB-202) tumor cell lines were cultured in RPMI-1640 supplemented with 10% FCS, 2 mM L-glutamine, 1 mM sodium pyruvate, 1x MEM NEAA and 50 µg/ml kanamycin (all from Bioconcept). The culture media for TF-1 cells was additionally supplemented with 10 ng/ml recombinant human GM-CSF (Peprotech #300-03). All human T-cell clones were maintained
820 in culture as previously described (72). A representative MAIT clone (MRC25) generated from blood of a healthy donor was previously characterized (72). Cells were free from Mycoplasma as evaluated by PCR analysis on DNA samples. When possible, cells were authenticated by staining with mAb for specific cell surface markers.

Lentiviral transductions were carried out as previously described (1) with plasmids indicate in
825 Table S2. Transduced cells were selected by FACS based on the expression of EGFP or mCherry reporters, or by 2 µg/mL puromycin (Calbiochem, Cat#540411) resistance.

Isolation and culture of primary cells from human blood and tissue samples

MR1T cells were isolated from the peripheral blood of healthy individuals as previously
830 described (72) with Lymphoprep (StemCell Technologies Cat#07851). In some experiments, purified T cells were stimulated with irradiated (80 Gray) A375-MR1 cells (ratio 2:1); 12 days later CD3⁺ M₃ADE-MR1-tetramer⁺ cells were sorted by FACS and cloned by limiting dilution in the presence of phyto-haemagglutinin (1 µg/mL, Remel, Cat#30852801 HA16), human rIL-2 (100 U/mL, Peprotech #200-02) and irradiated PBMCs (5×10⁵ cells/mL). T-cell clones were
835 periodically re-stimulated following the same protocol (72).

T cells, B cells, monocytes, myeloid dendritic cells (mDCs) and plasmacytoid dendritic cells (pDCs) were purified from PBMCs using immunomagnetic separation with specific kits from

StemCell Technologies (Cat#19051, 17954, 19359, 19061 and 19062A, respectively), according to manufacturer's protocol.

840 Tissue biopsy samples were derived from non-small cell lung cancer that were digested with media containing Accutase (Innovative Cell; Cat#AT-104), Collagenase IV 200 U/mL (Worthington; Cat#LS004189), DNase I 0.5 mg/mL (Sigma-Aldrich Cat#D5025) and Hyaluronidase 50mg/mL (Sigma; Cat#H6254) for 1 hour (h) at 37°C. Digested material was

845 stored in liquid nitrogen. After thawing, TILs were rested for 2 days prior to co-culture with A375-MR1 cells in a ratio of 1:1 and in the presence of 50 μ M M₃ADE. On day 5, human rIL-2 (5 U/mL) was added to the TIL cultures for a further 5 days, and expansion in this manner was repeated three times. The cells were then stained with MR1-M₃ADE tetramer, anti-CD3, anti-CD4 and anti-CD8 mAbs, and tetramer-positive cells were sorted into a bulk line before

850 functional experiments.

Human knockout library screening

A375-MR1-Cas9 cells generated using the previously described cell line (1) and Lenti Cas9-Blast plasmid (Addgene, Cat#52962) were transduced at 0.3 MOI by both part A and B of the

855 pooled Human GeCKO v2 CRISPR library (Addgene, Cat#1000000049), and subsequently selected using 2 μ g/mL puromycin (Calbiochem) for 96 h. Eight biological replicates of the resulting APCs, each with 64-fold representation of each guide within the library, underwent four consecutive rounds of killing by TC5A87 cells at a E:T of 2:1. The surviving cells were expanded for 24 h and DNA was extracted using a NucleoSpin Tissue kit (Macherey-Nagel,

860 Cat#740952). An additional eight biological replicates were similarly prepared but did not undergo killing by TC5A87, as to serve as controls. Illumina libraries were prepared as previously described (73). Briefly, JScispr1 and JScispr3 primers (Table S2) were used to amplify genomic sgRNA from the extracted gDNA and attach common Illumina primer handles for attaching sequencing library indexes. The former primer also inserts an 8-nt degenerate

865 sequence immediately downstream of the Illumina read 1 start site, decreasing issues of sequencing low-complexity libraries. Each replicate was barcoded by a unique pair of Nextera indexes (Illumina, Cat#15055290) in a second step PCR performed, as described in the Nextera DNA library preparation protocol (Illumina). A high-fidelity Advantage HF2 PCR kit (Takara, Cat#639123) was used in each of the PCR steps involved in preparing the sequencing libraries.

870 Libraries were quantified using a BioAnalyser high sensitivity DNA kit (Agilent, Cat#5067-4626) and a Qubit high-sensitivity dsDNA kit (ThermoFisher, Cat#Q32851) and pooled to form an

equimolar sequencing library that was denatured and diluted to 1.2 pM with a 20% PhiX v3 control library (Illumina, Cat#FC-110-3001), as described in the Illumina Denature and Dilution protocol (Illumina) before sequencing on a NextSeq500 using a High-output 150-cycle v2 kit
875 (Illumina, discontinued product). Both sets of sequencing libraries were sequenced using a dual-indexed single-end protocol (131 cycles on read 1, 8 cycles on each barcode) to a depth of 25 million reads per replicate, ensuring that guides depleted after T-cell mediated killing may be detected.

880 **CRISPR/Cas9-Mediated Gene Disruption**

Results obtained in the screening were confirmed by knock-out of selected genes in A375-MR1-Cas9 cells transduced with sgRNAs different from the ones present in the library (Table S2). After lentiviral transduction and selection, A375-MR1-Cas9 cells were maintained for a limited number of passages and used in activation assays as bulk populations. The expression
885 levels of target proteins were assessed by western blotting (Fig. S1) with antibodies against ADA (clone D1P4Y, Cell Signaling Technology, Cat#65184), ADSSL1 (clone 2D12, Sigma-Aldrich, Cat#SAB1401979), Glyoxalase 1 (clone 266, Novus Biologicals, Cat#NBP2-43618), LACC1 (clone E-12, Santa Cruz Biotechnologies, Cat#sc-376231) TPI1 (Proteintech, Cat#13937-1-AP), Actin (clone C-2, Santa Cruz Biotechnologies, Cat#sc-8432) or α -Tubulin (clone B-5-1-2, Sigma-Aldrich, Cat#T5168). Primary antibodies were revealed with IRDye[®] 680 Goat anti-Mouse IgG or IRDye[®] 800CW Goat anti-Rabbit IgG (LI-COR, Cat#926-32220 and Cat#926-32211, respectively) or HRP Goat anti-Mouse IgG (H+L) (ThermoFisher, Cat#A16072) and WesternBright Sirius (Advansta, Cat#K12043). MR1 surface expression of genetically-engineered cell lines was evaluated by flow cytometry with APC mouse anti MR1 mAbs 26.5
890 (Biolegend, Cat#361108), using APC Mouse IgG2a (clone MOPC-173, Biolegend, Cat#400220) as isotype control (Fig. S1).

The antigen presentation ability of different gene-edited cell lines was tested by stimulation of MAIT clone MRC25 after pulsing APCs for 2 h at 37°C with indicated concentrations of freshly-prepared 5-OP-RU (Fig. S1) as described in (74), with 5-A-RU (Toronto Research
900 Chemicals, Cat#A629245) and Methylglyoxal (MEG, Sigma-Aldrich, Cat#M0252).

Preparation and purification of synthetic antigens

Compounds M₃ADE, OPdA, M₁G, OPdC were synthesized and subsequently purified before use with cells.

905

Synthesis of 8-(9H-purin-6-yl)-2-oxa-8-azabicyclo[3.3.1]nona-3,6-diene-4,6-dicarbaldehyde (M₃ADE).

M₃ADE was synthesized as previously described (56) with some modifications. 1,1,3,3-Tetraethoxypropane (1.1 g, 5 mmol, 4.0 eq., Sigma-Aldrich, Cat#T9889) in aq. HCl (25 mL, 1 M) was stirred at 40°C for 1 h. Subsequently, a solution of adenine (168.9 mg, 1.25 mmol, 1.0 eq., Sigma-Aldrich, Cat#A8626) in water (25 mL) was added. The mixture was adjusted to pH 4.0 with aq. NaOH (1 M) and stirred for 5 days at 37 °C. M₃ADE was purified by solid phase extraction over Sep-Pak C18 2 g cartridges (Waters Corp., Milford, MA, Cat#WAT036915). Cartridges were preconditioned with 10 mL water and 10 mL Acetonitrile. Raw M₃ADE was washed with 20 mL water, 20 mL 10% Acetonitrile, then eluted with 20 mL 20% Acetonitrile. M₃ADE HPLC purification was performed on a JASCO RHPLC system equipped with an MD-4010 Photo Diode Array detector. Semi-preparative HPLC purification was performed using a 250 x 10 mm 5 μM NUCLEODUR C18 Pyramid HPLC column (Macherey-Nagel, Cat# 762272.100) at a temperature of 23°C where the mobile phases A and B were water and 95% methanol in water, respectively. Separation was performed with a flow rate of 6 mL/min with a linear gradient of 0-50% B from 0 to 15 min, 50-100% B from 15 to 38 min, 100% B from 38 to 43 min, 100-0% B from 43 to 44 min and 0% B until 50 min. M₃ADE yield was 12.5 mg (42 μmol, 3.4%). Biologically active HPLC peaks were collected for mass spectrometric and NMR analyses.

¹H-NMR (600 MHz, D₂O, δ/ppm): 9.24 (s, 1H, H₂₁), 9.11 (s, 1H, H₁₉), 9.09 (s, 1H, H₁₅), 8.61 (s, 1H, H₂), 8.38 (s, 1H, H₈), 7.64 (s, 1H, H₁₇), 7.33 – 7.29 (m, 1H, H₁₁), 4.09 – 4.07 (m, 1H, H₁₃), 2.21 (ddd, ²J_{H12a-H12b} = 13.7 Hz, ³J_{H12a-H11} = 2.8 Hz, ³J_{H12a-H13} = 2.8 Hz, 1H, H_{12a}), 2.03 (ddd, ²J_{H12b-H12a} = 13.7 Hz, ³J_{H12b-H11} = 2.2 Hz, ³J_{H12b-H13} = 2.2 Hz, 1H, H_{12b}).

¹³C-NMR (151 MHz, D₂O, extracted from HSQC and HMBC, δ/ppm): 193.2 (C₁₉), 192.2 (C₂₁), 166.5 (C₁₇), 154.3 (C₄), 152.7 (C₂), 150.3 (C₆), 149.3 (C₁₅), 144.8 (C₈), 126.2 (C₁₄), 125.7 (C₁₆), 121.8 (C₅), 79.9 (C₁₁), 25.0 (C₁₂), 17.4 (C₁₃).

HR-ESI-MS: calcd. for [M+Na]⁺ C₁₄H₁₁N₅NaO₃ m/z = 320.0754, found 320.0758.

Synthesis of Pyrimido[1,2-a]purin-10(3H)-one (M₁G, CAS 103408-45-3)

M₁G was synthesized as previously described (75) and (76) with some modifications. 1,1,3,3-Tetraethoxypropane (1.4 g, 6.25 mmol, 5.0 eq.) in aq. HCl (25 mL, 1 M) was stirred at 40°C for 1 h. Subsequently, a solution of guanine (188.9 mg, 1.25 mmol, 1.0 eq., Sigma-Aldrich, Cat#G11950) in aq. HCl (25 mL, 1 M) was slowly added. The mixture was stirred at 40 °C for 1 h and then kept at 4 °C for 16 h. The precipitate was washed three times with absolute ethanol

940 at 2000 x g for 10 min. The raw M₁G was extracted three times from the precipitate with 65°C water. The combined extracts were filtered with 0.22 µm filter. The mixture was adjusted to pH 7.0 with aq. NaOH (1 M).

M₁G HPLC purification was performed as for M₃ADE purification using the same type of HPLC column and mobile phases. Separation was performed with a flow rate of 6 mL/min and a
945 linear gradient of 0-20% B from 0 to 40 min, 20-100% B from 40 to 41 min, 100% B from 41 to 46 min, 100-0% B from 46 to 47 min and 0% B until 55 min. M₁G yield was 12.5 mg (66.8 µmol, 5.3%). Biologically active HPLC peaks were collected for mass spectrometric and NMR analyses.

¹H-NMR (600 MHz, D₂O, δ/ppm): 9.31 (d, ³J_{H13-H12} = 7.2 Hz, 1H, **H₁₃**), 8.97 (dd, ³J_{H11-H12} = 4.1 Hz,
950 ⁴J_{H11-H13} = 2.0 Hz, 1H, **H₁₁**), 8.22 (s, 1H, **H₈**), 7.30 (dd, ³J_{H12-H13} = 7.2 Hz, ³J_{H12-H11} = 4.2 Hz, 1H, **H₁₂**).

¹³C-NMR (151 MHz, D₂O, extracted from HSQC and HMBC, δ/ppm): 162.9 (**C₁₁**), 154.4 (**C₄**), 154.1 (**C₆**), 149.9 (**C₂**), 146.0 (**C₈**), 138.4 (**C₁₃**), 116.7 (**C₅**), 112.0 (**C₁₂**).

HR-ESI-MS: calcd. for [M+H]⁺ C₈H₆N₅O m/z = 188.0567, found 188.0571.

955 *Synthesis of N⁶-(3-Oxo-1-propenyl)-2'-deoxyadenosine (OPdA, CAS 178427-43-5)*

OPdA was synthesized as previously described (77) with some modifications. 2'-deoxyadenosine monohydrate (219 mg, 0.813 mmol, 1 eq., Sigma-Aldrich, Cat#D7400) was dissolved in 2 mL anhydrous dimethyl sulfoxide under argon atmosphere. Propargyl aldehyde (12 µL, 11.0 mg, 0.203 mmol, 0.25 eq., Toronto Research Chemicals, Cat#TRCP838440) was
960 added to the stirred solution and additional propargyl aldehyde (1.25 eq.) was added over a 72-h period. The reaction mixture was filtered and purified by preparative HPLC on a Shimadzu LC system (LC-20AT prominence liquid chromatograph, with an SPD-20A prominence UV/VIS detector (λ = 254 and 280 nm). Preparative HPLC purification was performed using a Reprisil-Pur 120 ODS 3,5 µM, 150 x 20 mm column (Maisch GmbH, Cat#r15.93.s1520), where the
965 mobile phases A and B were water and 90% acetonitrile in water, respectively. Separation was performed with a flow rate of 9 mL/min with a linear gradient of 1-30% B from 5 to 15 min, 30-100% B from 15 to 17 min, 100% B from 17 to 21 min, 100-0% B from 21 to 22 min and 1% B until 25 min. Analytical HPLC was performed with a LC-20AD prominence liquid chromatograph combined with a Shimadzu LCMS-2020 liquid chromatograph mass
970 spectrometer. Biologically active HPLC peaks were collected for mass spectrometric and NMR analyses. The OPdA yield was 13.5 mg (44.0 µmol, 5.4%).

¹H-NMR (500 MHz, D₂O, δ/ppm): 9.21 (d, ³J_{H13-H12} = 8.7 Hz, 1H, **H₁₃**), 8.49 (d, ³J_{H11-H12} = 13.5 Hz, 1H, **H₁₁**), 8.43 (s, 1H, **H₈**), 8.38 (s, 1H, **H₂**), 6.45 (dd, ³J_{H1'-H2'a} = 6.8 Hz, ³J_{H1'-H2'b} = 6.8 Hz, 1H, **H_{1'}**), 5.89 (dd, ³J_{H12-H11} = 13.5, ³J_{H12-H13} = 8.7 Hz, 1H, **H₁₂**), 4.65 (ddd, ³J_{H3'-H2'a} = 6.1 Hz, ³J_{H3'-H2'b} = 3.5 Hz,

975 $^3J_{H3'-H4'} = 3.5$ Hz, 1H, **H_{3'}**), 4.19 (ddd, $^3J_{H4'-H5'b} = 3.8$ Hz, $^3J_{H4'-H5'a} = 3.5$ Hz, $^3J_{H4'-H3'} = 3.5$ Hz, 1H, **H_{4'}**), 3.86 (dd, $^2J_{H5'a-H5'b} = 12.5$ Hz, $^3J_{H5'a-H4'} = 3.4$ Hz, 1H, **H_{5'a}**), 3.80 (dd, $^2J_{H5'b-H5'a} = 12.6$ Hz, $^3J_{H5'b-H4'} = 4.3$ Hz, 1H, **H_{5'b}**), 2.80 (ddd, $^2J_{H2'a-H2'b} = 13.7$ Hz, $^3J_{H2'a-H1'} = 7.1$ Hz, $^3J_{H2'a-H3'} = 6.4$ Hz, 1H, **H_{2'a}**), 2.58 (ddd, $^2J_{H2'b-H2'a} = 14.0$ Hz, $^3J_{H2'b-H1'} = 6.3$ Hz, $^3J_{H2'b-H3'} = 3.5$ Hz, 1H, **H_{2'b}**).

¹³C-NMR (126 MHz, D₂O, extracted from HSQC and HMBC, δ /ppm): 195.7 (**C₁₃**), 151.9 (**C₂**), 151.1 (**C₁₁**), 150.6 (**C₄**), 148.9 (**C₆**), 142.6 (**C₈**), 120.8 (**C₅**), 111.1 (**C₁₂**), 87.5 (**C_{4'}**), 84.7 (**C_{1'}**), 71.1 (**C_{3'}**), 61.6 (**C_{5'}**), 39.1 (**C_{2'}**).

HR-ESI-MS: calcd. for [M+Na]⁺ C₁₃H₁₅N₅NaO₄ m/z = 328.1016, found 328.1020.

Synthesis of N⁴-(3-Oxo-1-propenyl)-2'-deoxycytidine (OPdC, CAS 129124-79-4)

985 OPdC was synthesized as previously described (77) with some modifications. 2'-deoxycytidine (185 mg, 0.813 mmol, 1 eq., Sigma-Aldrich, Cat#D3897) was dissolved in 2 mL anhydrous dimethyl sulfoxide under argon atmosphere. Propargyl aldehyde (12.0 μ L, 11.0 mg, 0.203 mmol, 0.25 eq.) was added to the stirred solution and additional propargyl aldehyde (1.25 eq.) was added over a 72-h period.

990 HPLC purification was performed as described for OPdA. The OPdC yield was 7.00 mg (25.0 μ mol, 3.1%). Biologically active HPLC peaks were collected for mass spectrometric and NMR analyses.

¹H-NMR (500 MHz, D₂O, δ /ppm): 9.31 (d, $^3J_{H10-H9} = 8.6$ Hz, 1H, **H₁₀**), 8.29 (d, $^3J_{H8-H9} = 13.7$ Hz, 1H, **H₈**), 8.19 (d, $^3J_{H6-H5} = 7.4$ Hz, 1H, **H₆**), 6.28 (d, $^3J_{H5-H6} = 7.4$ Hz, 1H, **H₅**), 6.24 (dd, $^3J_{H1'-H2'a} = 6.1$ Hz, $^3J_{H1'-H2'b} = 6.1$ Hz, 1H, **H_{1'}**), 5.91 (dd, $^3J_{H9-H8} = 13.7$ Hz, $^3J_{H9-H10} = 8.6$ Hz, 1H, **H₉**), 4.43 (ddd, $^3J_{H3'-H2'a} = 6.4$ Hz, $^3J_{H3'-H2'b} = 4.3$ Hz, $^3J_{H3'-H4'} = 4.3$ Hz, 1H, **H_{3'}**), 4.12 (ddd, $^3J_{H4'-H5'b} = 4.8$ Hz, $^3J_{H4'-H5'a} = 4.1$ Hz, $^3J_{H4'-H3'} = 4.1$ Hz, 1H, **H_{4'}**), 3.87 (dd, $^2J_{H5'a-H5'b} = 12.5$ Hz, $^3J_{H5'a-H4'} = 3.5$ Hz, 1H, **H_{5'a}**), 3.77 (dd, $^2J_{H5'b-H5'a} = 12.5$ Hz, $^3J_{H5'b-H4'} = 5.3$ Hz, 1H, **H_{5'b}**), 2.55 (ddd, $^2J_{H2'b-H2'a} = 14.1$ Hz, $^3J_{H2'b-H1'} = 6.3$ Hz, $^3J_{H2'b-H3'} = 4.3$ Hz, 1H, **H_{2'b}**), 2.32 (ddd, $^2J_{H2'a-H2'b} = 14.2$ Hz, $^3J_{H2'a-H1'} = 6.5$ Hz, $^3J_{H2'a-H3'} = 6.5$ Hz, 1H, **H_{2'a}**).

¹³C-NMR (126 MHz, D₂O, extracted from HSQC and HMBC, δ /ppm): 196.2 (**C₁₀**), 161.5 (**C₄**), 156.8 (**C₂**), 149.6 (**C₈**), 144.3 (**C₆**), 112.1 (**C₉**), 96.8 (**C₅**), 87.1 (**C_{1'}**), 87.1 (**C_{4'}**), 70.3 (**C_{3'}**), 61.1 (**C_{5'}**), 39.8 (**C_{2'}**).

HR-ESI-MS: calcd. for [M+Na]⁺ C₁₂H₁₅N₃NaO₅ m/z = 304.0904, found 304.0902.

1005

MS and NMR analysis

Unless otherwise stated, chemicals were used as received without further purification. NMR analysis of all of the antigens was performed at 298 K on a Bruker Avance III NMR spectrometer operating at 500 MHz proton frequency equipped with a BBFO probehead or

1010 on a Bruker Avance III HD NMR spectrometer operating at 600 MHz proton frequency
equipped with a cryogenic QCI-F probe. Standard pulse sequences were used for cosy, tocsy,
noesy, hsqc, hmqc and hmbc 2D-NMR experiments and the spectra were processed using the
topspin 4.0 software package. All compounds were fully characterized by means of 2D-NMR
and HR-ESI-MS. For all compounds ¹H- and ¹H-¹³C-HSQC or ¹H-¹³C-HMQC spectra, as well as
1015 experimental and calculated HR-ESI-MS spectra are shown in Figure S3 and S4.
HR-ESI-MS spectra were measured on a Bruker MaXis 4G high resolution ESI Mass
Spectrometer in direct injection mode using methanol containing 0.1% v/v formic acid.

Evaluation of compound biological activity

1020 Cell surface MR1 upregulation

THP-1 cells (10⁵ cells/well) were tested for MR1 surface expression after incubation with or
without synthetic compounds: M₃ADE (1 μM), OPdA (100 μM), M₁G (13μM) and OPdC (100
μM) for 6 h at 37°C. Ac-6-FP (100 μM) (Schircks Laboratories, Cat# 11.418) was used as positive
control for MR1 surface upregulation. The cells were stained with an anti-human-MR1-APC
1025 mAb (clone 26.5, Biolegend) or with APC-labelled mouse IgG2a, k isotype control (clone
MOPC-173, Biolegend) antibodies for 20 min at 4°C, then washed and analyzed by flow
cytometry. For each condition, net MFI was calculated subtracting isotype MFI from anti-MR1
MFI and fold change of cells treated with synthetic molecules over cells treated with vehicle
was calculated.

1030

T-cell activation assays

Activation assay with living or fixed APCs

MR1T cells (5×10⁴/well unless otherwise indicated) were co-cultured with the indicated APCs
(10⁵ cells per well unless otherwise indicated) for 18 h in 120 μL volume in triplicate. In some
1035 experiments, anti-MR1 mAbs (Ultra-LEAF™ Purified clone 26.5, Biolegend, Cat#361110) or
mouse IgG2a isotype control mAbs (Ultra-LEAF™ Purified clone MOPC-173, Biolegend,
Cat#400263) (both at 30 μg/mL) were added and incubated for 30 min at 37°C prior to the
addition of T cells. In certain experiments, Ultra-LEAF™ purified mouse mAb anti-human-HLA-
A,B,C (clone W6/32) and Ultra-LEAF™ purified anti-human HLA-DR Antibody (clone L243) (20
1040 μg/mL each, Biolegend, Cat#311428 and #307648, respectively) were added to both anti MR1
or Isotype control antibodies.

When nucleobases, nucleosides or nucleotides (all 250 μM, all from Sigma-Aldrich) and
synthetic compounds M₃ADE, OPdA, M₁G or OPdC were used to stimulate T cells, the THP-1

cells (10^5 /well) were cultured 2 h with the indicated molecules or medium only, prior to T-cell
1045 addition. When a single concentration of synthetic antigens was used, 100 μ M OPdA, 100 μ M
OPdC and 13 μ M M₁G were used for all clones, while 100 μ M M₃ADE was used for all clones
except DGB129, AC1A4, AC1B76 and AVA46 for which 1 μ M was used.

In experiments with Mycophenolic acid (10 μ M, Sigma-Aldrich, Cat#M3536), EHNA (25 μ M,
Sigma-Aldrich, Cat#E114), S-p-bromobenzylglutathione cyclopentyl diester (BBG, 20 μ M,
1050 Sigma-Aldrich, Cat#SML1306), THP-1 cells (1×10^6 /mL) were treated with the indicated
concentrations of drugs in complete medium at 37°C for 18 h before being washed twice with
PBS, counted and used for T-cell activation. In experiments where THP-1 cells (1×10^6 /mL) were
treated with Doxorubicin (75 nM, Selleck Chemical, Cat#S1208) and Paclitaxel (5 μ M, Sigma-
Aldrich, Cat#T7191), the cells were incubated at 37°C for 18 h before being washed twice with
1055 PBS, counted, and incubated for 2 h with vehicle or 150 μ M dAdo and 150 μ M Guanosine
(Sigma-Aldrich, Cat#D7400 and G6752, respectively) prior to T-cell addition. In experiments
where fixed A375-MR1 cells were used to activate MR1T cells, APCs (4×10^5 /mL) were treated
with Apocynin (APO, 100 μ M, Sigma-Aldrich, Cat#W508454), L-Glutathione reduced (GSH, 4
mM, Sigma-Aldrich Cat#G6013), N-acetyl cysteine (NAC, 4 mM, Sigma-Aldrich, Cat#A9165), L-
1060 Buthionine-sulfoximine (BSO, 400 μ M, Sigma-Aldrich, Cat#B2515) Mercaptosuccinic acid
(MSA, 3.3 μ M, Sigma-Aldrich, Cat#M6182), ML-210 (6 μ M, Sigma-Aldrich Cat#SML0521) or
1S,3R-RSL3 (RSL3, 1 μ M, Sigma-Aldrich, Cat#SML2234), Hydralazine hydrochloride (100 μ M,
Sigma-Aldrich, Cat#H1753) or Aminoguanidine hemisulfate salt (5 mM, Sigma-Aldrich,
Cat#A7009) for 18 h at 37°C before being washed twice with PBS, fixed with 0,05%
1065 Glutaraldehyde (Sigma-Aldrich, Cat#G5882), counted and used for MR1T cell stimulation.
In some experiments, MAIT cells were stimulated by APCs pulsed 3 h with 5-OP-RU as
previously described (74).

To confirm that drugs reducing MR1T stimulation do not affect MR1 presentation ability,
A375-MR1 cells treated with different molecules were collected before fixation and used to
1070 stimulate the MAIT clone MRC25 after pulsing for 2 h at 37°C with the indicated
concentrations of freshly-prepared 5-OP-RU (74) or with 30 μ M 6,7-dimethyl-8-
Ribityllumazine (Cayman Chemical, Cat#23370).

Activation assay with plate-bound soluble MR1

1075 Recombinant human β 2m-MR1-Fc was produced in CHO-K1 cells as previously described (1)
and 4 μ g/mL were coated onto 96 wells plates (Nunc, Cat#439454) for 18 h at 4°C. Plate-bound
MR1 was then washed twice with wash buffer (150 mM NaCl, 20 mM Tris and 2% Glycerol, pH

5.6) to remove bound antigens. Then, the synthetic antigens (M₃ADE, OPdA, M₁G and OPdC) were added at the indicated concentrations and incubated for 6 h at room temperature (RT).
1080 Unbound antigens were washed twice with PBS before the addition of excess PBS. In some experiments, bacteria-produced and refolded MR1-M₃ADE protein was serially diluted in PBS and added to a high protein binding plate (Nunc) for 2 h at 37°C then washed twice and used in the stimulation assay. Indicated MR1T cell clones (10⁵/100 μl/well) were added and supernatants were collected after 18 h. Released cytokines were detected by ELISA.

1085

Cytokine Analysis

The following human cytokines were assessed by ELISA as previously described (1): human IFN-γ (capture MD-1 mAb; revealing biotinylated 4S.B3 mAb, Biolegend Cat#507502 and 502504, respectively), human IL-13 (capture JES10-5A2 mAb; revealing biotinylated SB126d
1090 mAb, SouthernBiotech Cat#10125-01 and 15930-08, respectively).

MR1 protein production and tetramerization

Soluble recombinant MR1 monomers were generated as previously described (7). Briefly, nucleotide sequences coding for the soluble portions of the mature human MR1 (GenBank
1095 accession number NM_001531) and human β2m (GenBank accession number NM_004048.4) were cloned into the bacterial expression vector pET23d (Novagen, Cat#69748-3). Transformed *E. coli* BL21(DE3)pLysS (Promega, Cat#L1195) were then grown to OD_{600nm} 0.4-0.6 before induction with 0.6M isopropyl β-D-1-thiogalactopyranoside (Sigma-Aldrich, Cat#10724815001). After 4 h of further culture, the cells were lysed and inclusion bodies were
1100 cleaned, purified and fully denatured in 8 M urea, 10 mM EDTA, 0.1 mM DTT for subsequent storage at -80°C.

For protein refolding, MR1 heavy chain (4mM), β2m (2mM) and ligand (15mM) were added to 1L refolding buffer containing 0.4M L-arginine (Alfa Aesar, Cat# A15738), 100mM Tris pH 8.0, 2 mM EDTA that was cooled to 4°C and supplemented with 5 mM reduced glutathione
1105 (Merk Millipore, Cat# G4251) and 0.5 mM oxidized glutathione (Merk Millipore, Cat# G4376) immediately beforehand. After 3 days, the refold mix was concentrated down to 10 mL and the refolded MR1 bearing ligand was purified by HPLC using sequential Sephacryl S300 16/60 (GE Healthcare, Cat#17116701), MonoQ 5/50 (GE Healthcare, Cat#17516601) chromatography (Figure S6A).

1110 The correct protein conformation was confirmed by performing a plate-bound activation of the MR1T cell clone DGB129, as described above. The functional monomer was then

biotinylated using BirA-500 biotinylation kit (Avidity, Cat# Bulk BirA) overnight at 4°C. Excess biotin was removed using S75 10/30 (GE Healthcare, 29148721) gel filtration before tetramerization with phycoerythrin (PE)-streptavidin (Prozyme, Cat# PJRS25) in a 4:1 molar ratio.

Immunofluorescence Staining

Cell surface labeling was performed using standard protocols, with mouse mAb anti-human-CD3-BV421 (clone UCHT1, Biolegend Cat#300434), mouse mAb anti-human-CD3-BV711 (clone OKT3, Biolegend Cat#317328), mouse mAb anti-human-CD4-APC (clone OKT4, Biolegend Cat#317416), mouse mAb anti-human-CD8a-BV711 (clone RPA-T8, Biolegend Cat#301044), mouse mAb anti-human-CD45RA-AF700 (clone 2H4LDH11LDB9 (2H4), Beckman Coulter Cat#B90396), mouse mAb anti-human-CCR7-AF488 (clone G043H7, Biolegend Cat#353206), or mouse mAb anti-human-HLA-A,B,C-APC (clone W6/32, Biolegend Cat#311410). All mAbs for staining were titrated on appropriate cells before use.

When staining with MR1-M₃ADE-, MR1-6-FP- (NIH Tetramer Core Facility #33024) or MR1-5-OP-RU-tetramers (NIH Tetramer Core Facility #33021), the cells were pre-treated with 50 nM dasatinib (Sigma-Aldrich, Cat# CDS023389) for 30 min at 37°C. In some experiments, mouse anti-human-Vβ8-APC (clone JR.2, Biolegend Cat#348106) or anti-CD8 mAbs were added for 20 min at RT, then 2.5 µg/mL tetramer for a further 20 min at RT without washing. All remaining mAbs were then added for a further 20 min at RT without washing. The cells were then washed in cold PBS before acquisition. The samples were acquired on an LSR Fortessa flow cytometer with the FACS Diva software (Becton Dickinson) or on CytoFLEX with CytExpert software (Beckman Coulter). Cell sorting experiments were performed using either Influx or FACSaria (Becton Dickinson). Dead cells and doublets were excluded on the basis of forward scatter area and width, side scatter, and DAPI (Sigma Cat#D9542) or Live/Dead DRAQ7™ (Thermo Fisher Scientific Cat#424001) staining, as depicted in Figure S6H. When PBMCs were analyzed, CD14⁺ and CD19⁺ cells were excluded by staining with fluorescent mouse mAb anti-human-CD14-APCCy7 (clone 63D3, Biolegend Cat#367108), mouse mAb anti-human-CD19-APCCy7 (clone HIB19, Biolegend Cat#302218), detected in the same channel as the Live/Dead staining. All data were analyzed using FlowJo (v10.5.3, Becton Dickinson).

ROS production measurement

CM-H₂DCFDA (Thermo Fisher Scientific Cat#C6827) was used to assess ROS production in cell upon cell treatment with Doxorubicin and Paclitaxel. THP-1 cells (10⁷/mL) were labelled with

10 μ M CM-H₂DCFDA for 30 min at 37°C in the dark, then washed with PBS and resuspended in complete medium. 10⁵ cells were seeded per well and treated with 75 nM Doxorubicin, 5 μ M Paclitaxel or vehicle for 18 h at 37°C. Phorbol 12-myristate 13-acetate (PMA, 50 ng/mL, Sigma-Aldrich Cat#P8139) was used as positive control.

1150

sgRNA-Seq Data Processing and analysis

Raw sequencing data (deposited at NCBI GEO #GSE160366) was demultiplexed using bcl2fastq (Illumina, v2.17.1.14) and read quality checked using FastQC (Babraham Bioinformatics, v0.11.4). Reads were then trimmed to remove the homologous regions flanking the sgRNA sequences using Trimmomatic v0.36 (78) using options HEADCROP:42 CROP:20. These trimmed reads were again passed through FastQC to check that average phred33 quality in the sgRNA sequences was >30. These reads were aligned to a GeCKO v2 sgRNA reference index using Bowtie2 (v2.2.9) (79) with options --very-sensitive-local. Read counts were then extracted from the resulting SAM files using custom perl script map_count.pl (Cox, M. available on request) and imported into R (80) for analysis using edgeR (81).

1155

Inequality of sgRNA activity in the GeCKO library (82) hinders hit-selection using rank-based methodology, thus differential enrichment analysis was performed with the edgeR package (81) using the GLM Robust method to estimate dispersions, after removing guides targeting known essential genes (16). The level of random enrichment and depletion of guides was estimated using the log₂ fold-change in the top and bottom 1% of negative control guides in the GeCKO library. Thus, guides with an FDR < 0.05 and log₂ fold-change greater or lower than the top and bottom 1% of negative control guides, respectively, were said to be significantly enriched or depleted by our screen. GO-term and KEGG-pathway enrichment analysis was performed using a binomial test on the significant unique gene-targets identified in the differential enrichment analysis (GO.db, v3.7.0)(18, 19, 83). Genes were annotated using biomaRt version 2.42.0 (84).

1160

1165

1170

Metabolic Modeling

To perform structural sensitivity analysis (21, 85), the Recon3D model was pre-processed by making all reactions reversible and removing biomass, maintenance, sink, and demand reactions. For each gene in the model, associated reactions were aligned by reversing reaction directions such that the overlap of metabolites on each side of the reaction equations was maximal. Assuming that perturbation of a gene implied identical flux perturbations for all associated reactions, flux balance analysis (FBA) (86) was used to check if this identical flux

1175

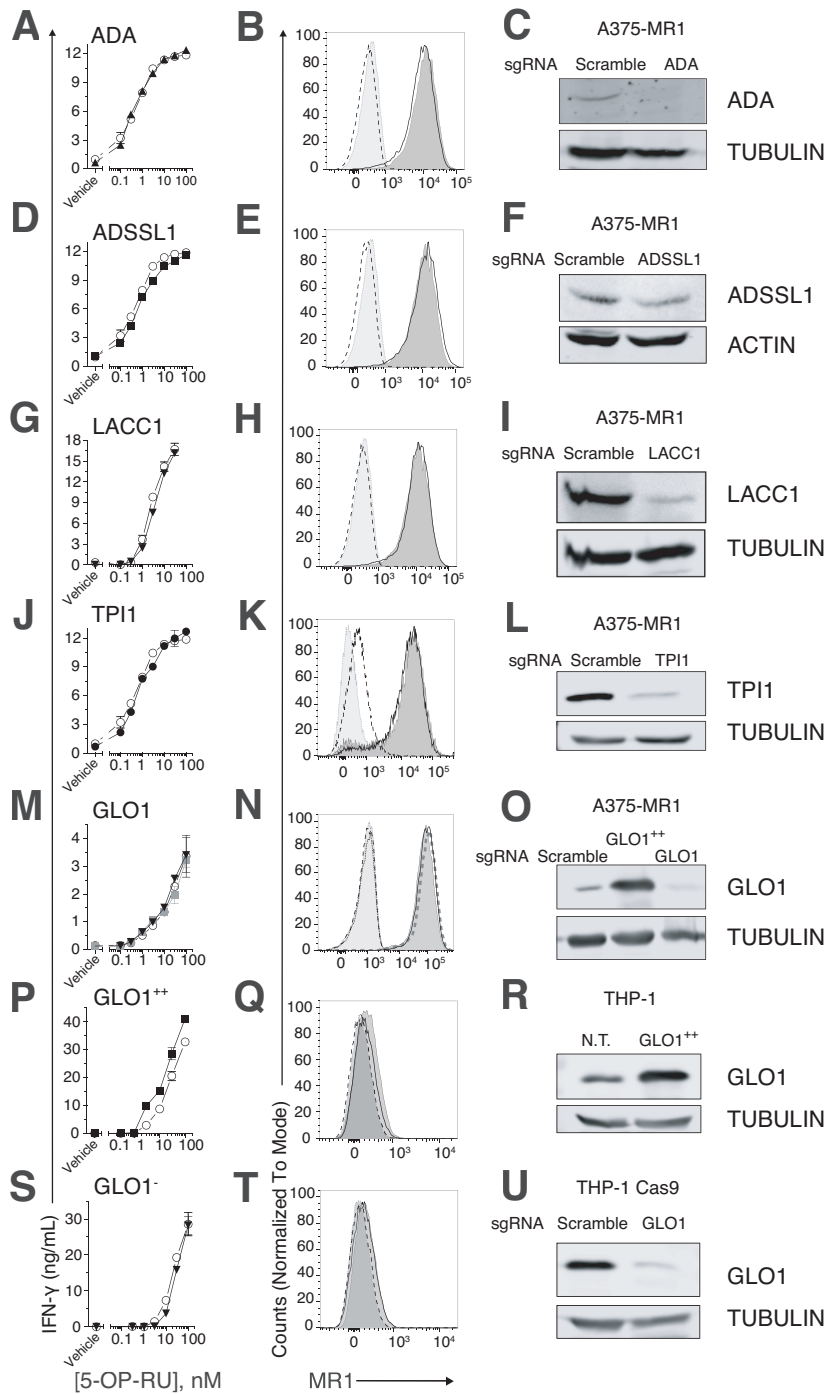
1180 perturbation was feasible at steady state. If it was infeasible, the closest feasible perturbation
was found by minimizing the Euclidean distance to the identical flux perturbation. Finally,
perturbed fluxes were fixed and the minimal network response was computed by minimizing
the Euclidean norm of all fluxes not directly linked to the perturbed gene. If the minimal
1185 response was a thermodynamically infeasible loop (87), constraints and variables were added
to specifically disable this loop (88) and a minimal response was re-computed until the
minimal loopless response was found. Genome-wide knock-out sensitivities were analyzed in
R to identify sets of >2 knock-outs with Pearson correlation scores > 0.6.

Data Analysis and Statistical Analysis

1190 Data analysis, statistical tests and visualization were conducted in R and GraphPad Prism. In
Prism (v8.1.0, GraphPad Software, Inc.), analysis was performed using multiple t-test, One- or
Two-way Anova as indicated for each assay in figure legends.
A p value < 0.05 was considered statistically significant. *p < 0.05, **p ≤ 0.01, ***p ≤ 0.005.

1195 **Data availability**

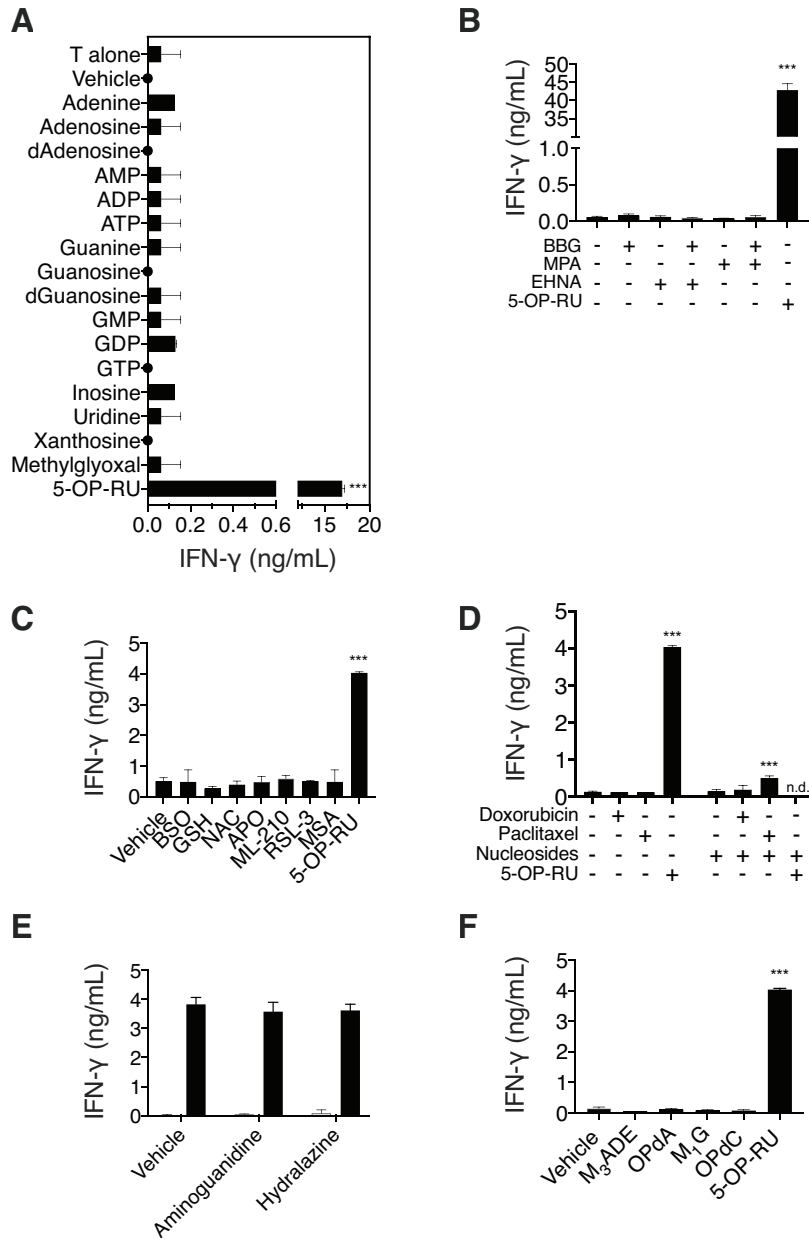
sgRNA sequencing data have been deposited to the NCBI GEO repository accession
GSE160366.



1200 **Fig. S1. Genetically-engineered cell lines characterization.**

(A, D, G, J, M, P, S) Activation assay of MAIT clone MRC25 in response to A375-MR1 (A, D, G, J, M) and THP-1 (P and S) cells pulsed with 5-OP-RU. Cells are either wild type (○), knock-out (A ▲, D ■, G ▼, J ●, M ▼, S ▼) or overexpressing (M ■, P ■) the indicated genes. IFN- γ is expressed as mean \pm SD of triplicate independent cultures.

- 1205 **(B, E, H, K, N, Q, T)** Surface MR1 expression of the genetic engineered cell lines. MR1 staining of DAPI-negative wild type (dark grey shadow), knock-out (black line) and GLO1-overexpressing (**N**, grey thick dashed line) cells with anti-MR1 mAbs 26.5. Isotype-matched control staining is depicted in wild type (light grey shadow with grey dot line), in knock-out (black dashed line) or GLO1-overexpressing (**N**, black dotted line) cells.
- 1210 **(C, F, I, L, O, R, U)** Western blot analysis of target protein expression in indicated cell lines. Tubulin or Actin were used as loading control.
- The experiments were repeated at least twice and one representative experiment is shown.



1215

Fig. S2. Stimulation of MAIT clone MRC25 with nucleobases, inhibitory drugs and synthetic antigens.

(A) MAIT clone MRC25 stimulated with THP-1 cells in the presence of the indicated nucleobases (250 μM), Methylglyoxal (250 μM) or 5-OP-RU (30 nM).

1220

(B and D) MRC25 cells stimulated with THP-1 cells treated with indicated drugs.

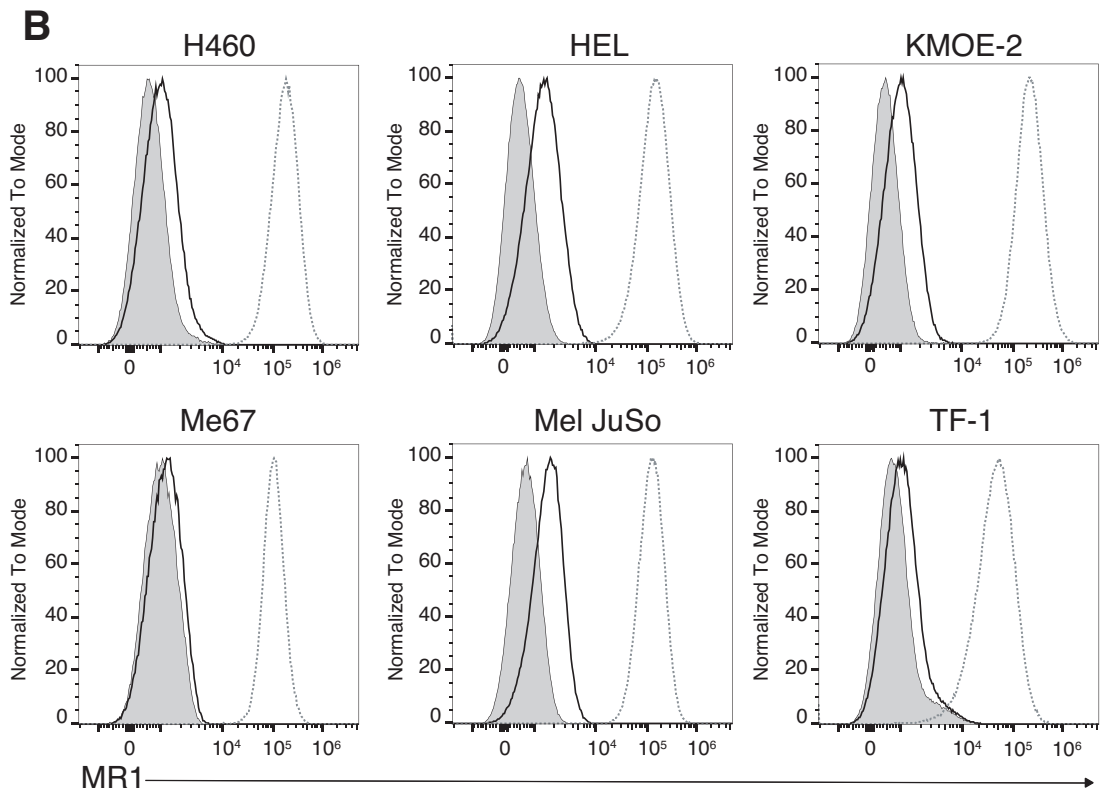
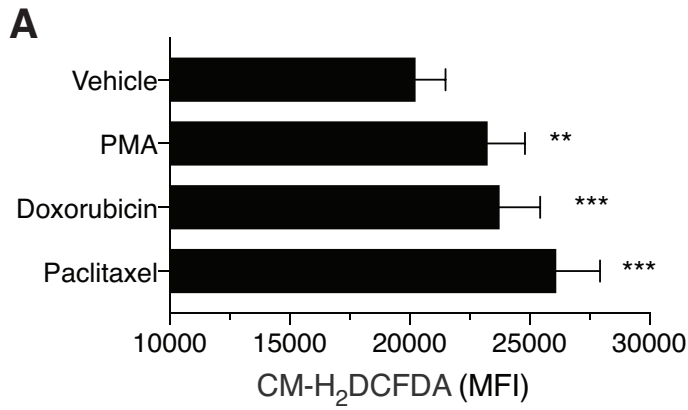
(C) MRC25 cells stimulated with A375-MR1 cells treated with GSH, NAC, APO, BSO or GPX inhibitors and fixed or with THP-1 cells pulsed with 5-OP-RU (10 nM).

(E) MRC25 cells stimulated with A375-MR1 cells treated with carbonyl scavengers and fixed before T cell addition (empty bars). As controls, the same experiment was performed with the same carbonyl scavengers in the presence of 6,7-dimethyl-8-Ribityllumazine (20 μ M, black bars).

(F) MRC25 cells stimulated with THP-1 cells in the presence of M₃ADE, OPdA, OPdC (all 100 μ M), M₁G (13 μ M) or 5-OP-RU (10 nM).

n.d., not determined. IFN- γ is expressed as mean \pm SD of triplicate independent cultures. The experiments were repeated at least twice and one representative experiment is shown.

*** $p \leq 0.001$ compared to vehicle-treated cells using One-way ANOVA (A, B, C, F) or Two-way ANOVA (D and E) with Dunnett's multiple comparison.

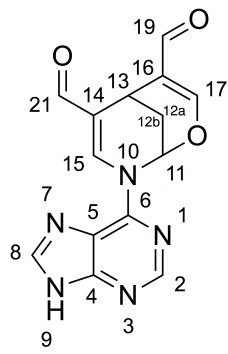
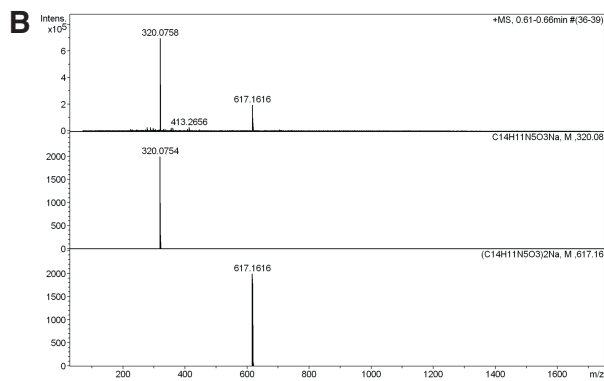
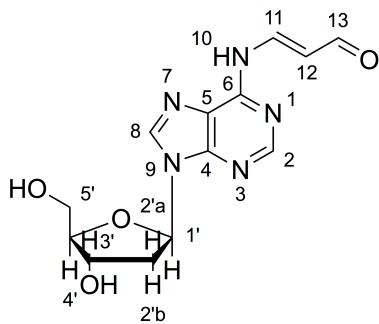
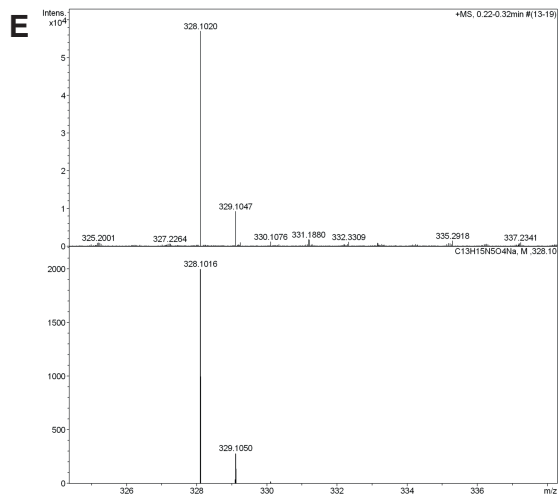
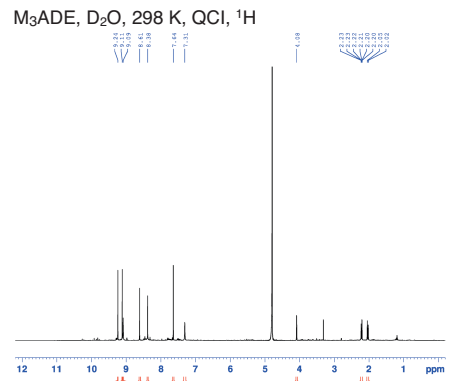
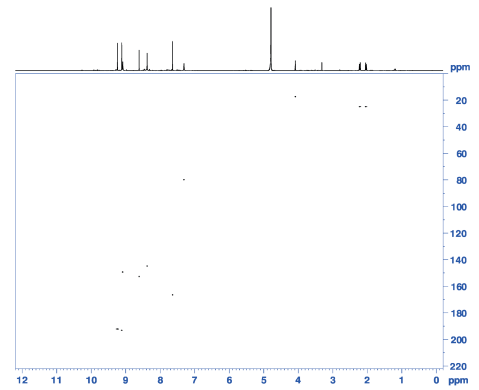
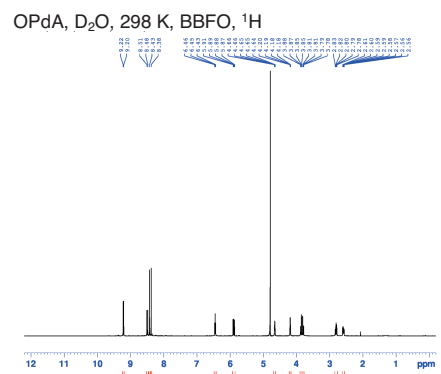
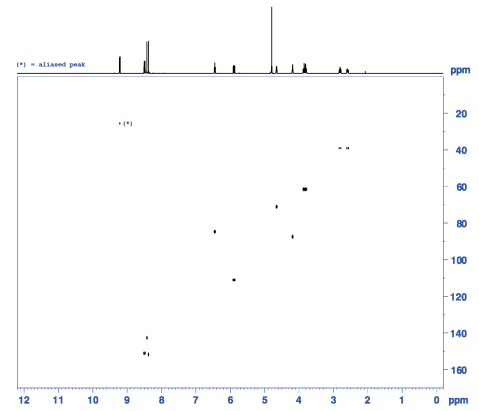


C

Cell line	Tissue origin	Disease
H460	Lung	Large cell lung cancer
HEL	Bone Marrow	Erythroleukemia
KMOE-2	Bone Marrow	Erythroleukemia
Me67	Skin	Melanoma
Mel JuSo	Skin	Melanoma
TF-1	Bone Marrow	Erythroleukemia

Fig. S3. ROS accumulation in APCs and MR1 expression in tumor cell lines.

- 1240 (A) Quantification of ROS produced in THP-1 cells treated with Doxorubicin, Paclitaxel or Phorbol 12-myristate 13-acetate (PMA). Results are expressed Median Fluorescence Intensity (MFI) of live cells \pm SD of triplicate independent cultures.
- (B) Surface expression in indicated tumor cell lines of MR1 (black line) or HLA A,B,C (grey dotted line). Isotype matching staining control is depicted as grey shade.
- (C) Table reporting tissue origin and diagnosed disease of each cell line.
- Each experiment was repeated at least twice and one representative experiment is shown.
- 1245 ** $p \leq 0.01$ and *** $p \leq 0.001$ using one-way Anova with Dunnett's multiple comparison.

A**B****D****E****C**M₃ADE, D₂O, 298 K, QCI, hsqc**F**OPdA, D₂O, 298 K, BBFO, hmqc

1250 **Fig. S4. Mass-Spectrometric and NMR analyses of synthetic M₃ADE and OPdA.**
(A) 8-(9*H*-purin-6-yl)-2-oxa-8-azabicyclo[3.3.1]nona-3,6-diene-4,6-dicarbaldehyde (M₃ADE) structure,
(B) High Resolution Mass-Spectrometry (HRMS), top: experiment, center & bottom: calculated, and
1255 (C) NMR analysis. (D) N⁶-(3-oxo-1-propenyl)-2'-deoxyadenosine (OPdA) structure, (E) HRMS, top: experiment, bottom: calculated, and (F) NMR analysis.

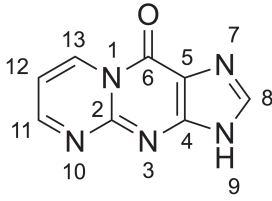
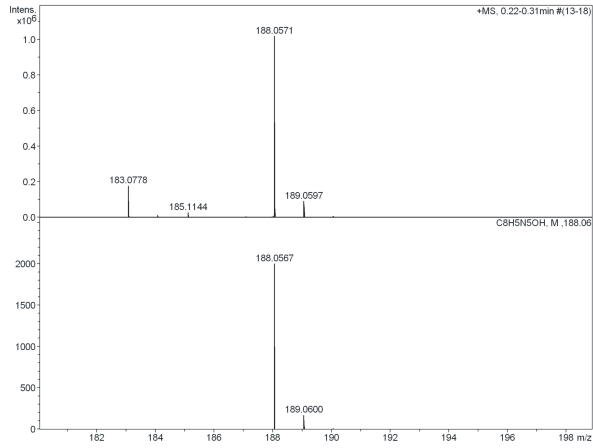
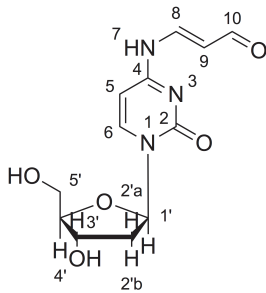
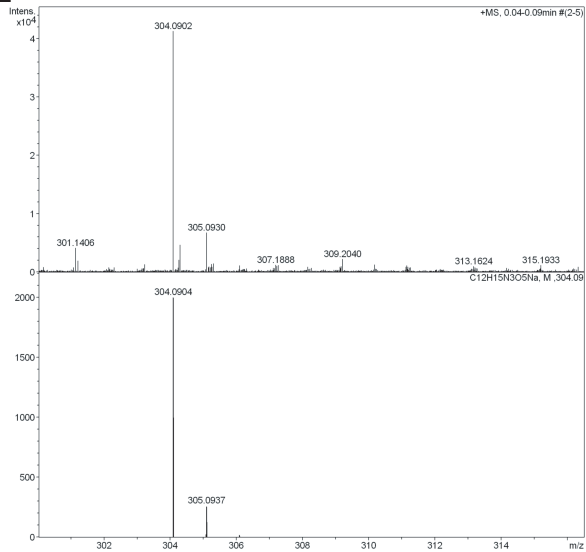
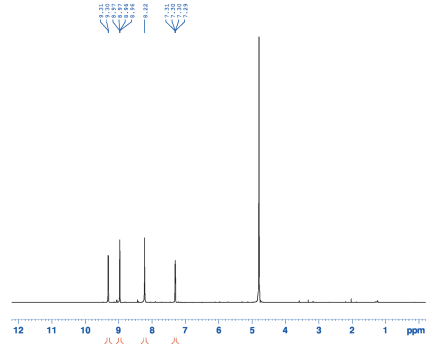
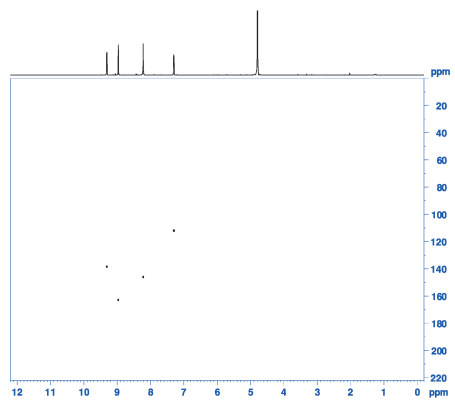
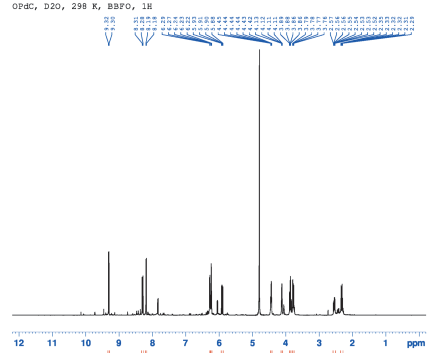
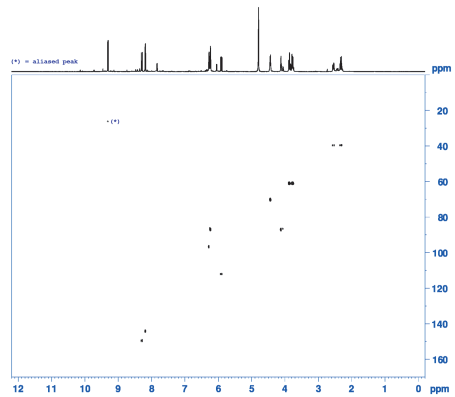
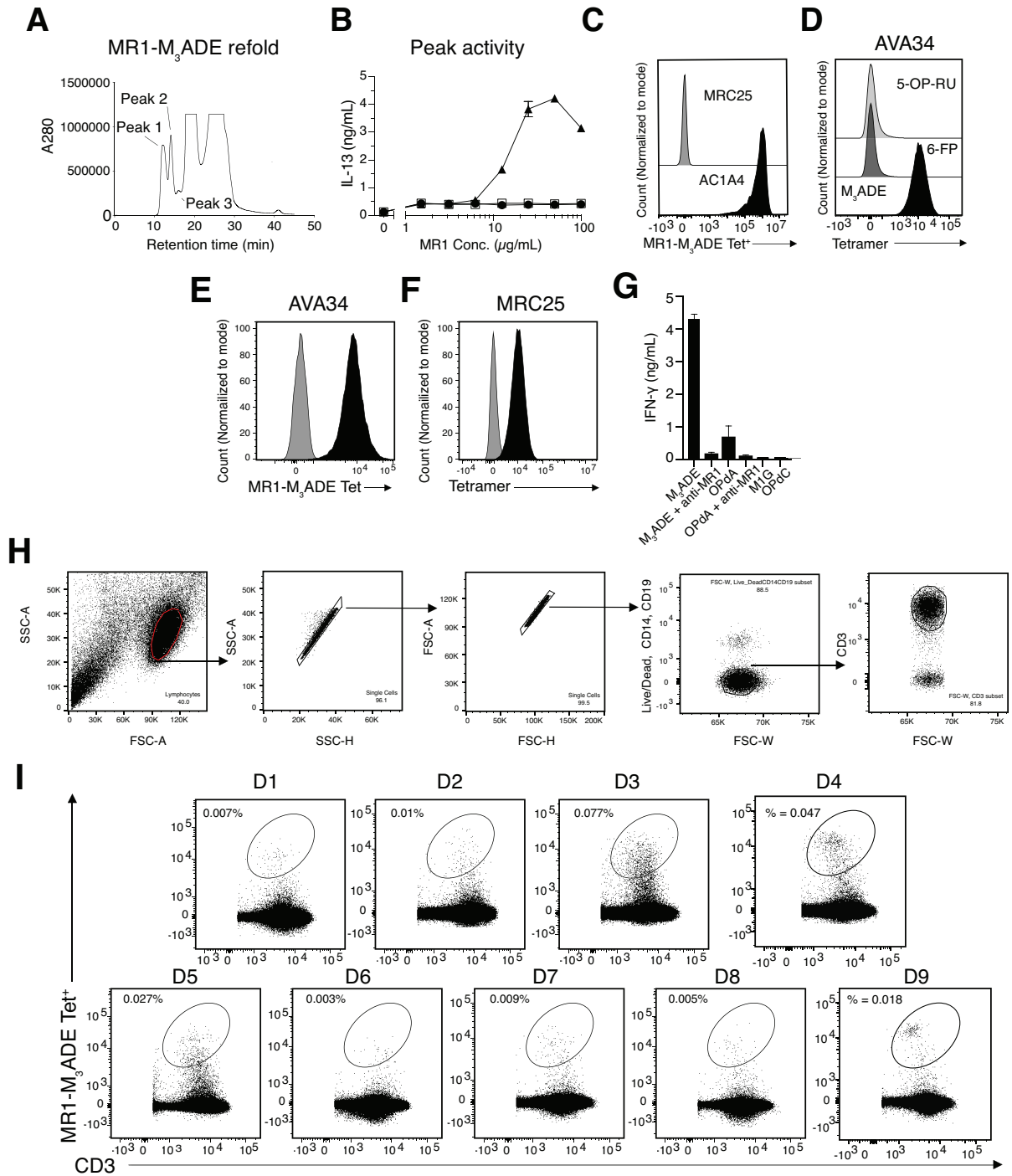
A**B****D****E****C** M₁G, D₂O, 298 K, QCI, ¹HM₁G, D₂O, 298 K, QCI, hsqc**F** OPdC, D₂O, 298 K, BBFO, ¹HOPdC, D₂O, 298 K, BBFO, hmqc

Fig. S5. Mass-Spectrometric and NMR analyses of synthetic M1G and OPdC.

(A) pyrimido[1,2- α]purin-10(3*H*)-one (M₁G) structure, (B) HRMS, top: experiment, bottom: calculated, and (C) NMR analysis.

1265 (D) *N*⁴-(3-oxo-1-propenyl)-2'-deoxycytidine (OPdC) structure, (E) HRMS, top: experiment, bottom: calculated, and (F) NMR analysis.



1270

Fig. S6. Generation of MR1-M₃ADE tetramer.

(A) Size exclusion chromatography of MR1 protein refolded in the presence of M₃ADE antigen. The indicated peaks 1, 2 and 3 were fractionated.

(B) Plate-bound protein from peaks 1 (square), 2 (triangle) and 3 (circle) were used to activate

1275 the clone DGB129 in a dose response manner and IL-13 was measured by ELISA.

(C) Histograms of MR1-M₃ADE tetramer staining of the MR1T-cell clone AC1A4 (black) compared to the MAIT-cell clone MRC25 (grey). T cell clone staining results are representative of at least two independent experiments.

1280 (D) Histograms of AVA34 cell staining (a representative clone derived from MR1-M₃ADE tetramer-sorting) using MR1 tetramers loaded with 5-OP-RU (light grey), 6-FP (dark grey) and M₃ADE (black). Results are representative of at least two independent experiments.

(E) Staining of the clone AVA34 with MR1-M₃ADE tetramer (Black) compared with AVA34 incubation with an anti-V β 8 antibody prior to MR1-M₃ADE tetramer staining (Grey).

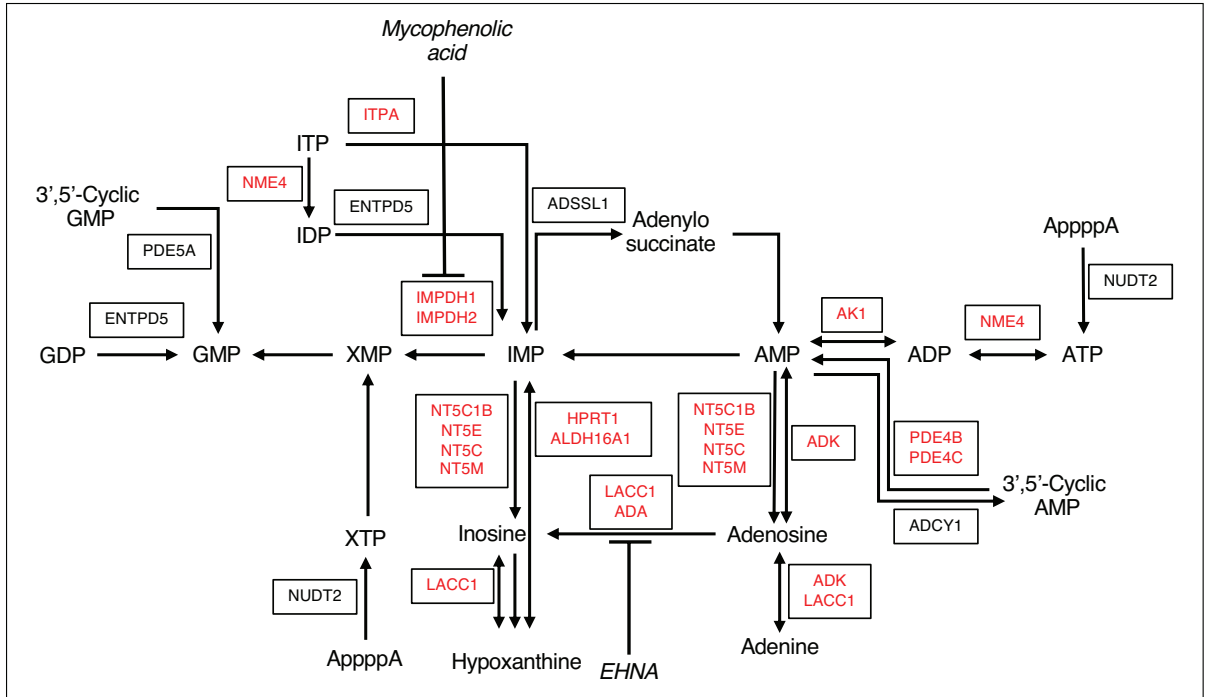
1285 (F) Staining of the MAIT clone MRC25 with either the MR1-M₃ADE tetramer (Grey) or the MR1-5-OP-RU tetramer (Black).

(G) Activation of clone AVA34 with 5 synthetic adducts, including relevant blocking with anti-MR1 mAbs. Columns show IFN- γ release (mean \pm SD of triplicate cultures). The activation data shown are representative of at least three independent experiments.

1290 (H) Flow cytometry gating strategy. Lymphocytes were gated among acquired events by FSC-A/SSC-A, doublets exclusion performed by selection with SSC-H/SSC-A and subsequently FSC-H/FSC-A to obtain singlets among which Live and CD14/CD19 double negative cells were selected prior identification of CD3⁺ T cells. One representative donor is shown.

1295 (I) MR1-M₃ADE tetramer staining of PBMCs from nine healthy donors plotted against CD3 expression. Cells are pre-gated on live, single cells and the frequency is presented as a percentage of CD3⁺ T cells.

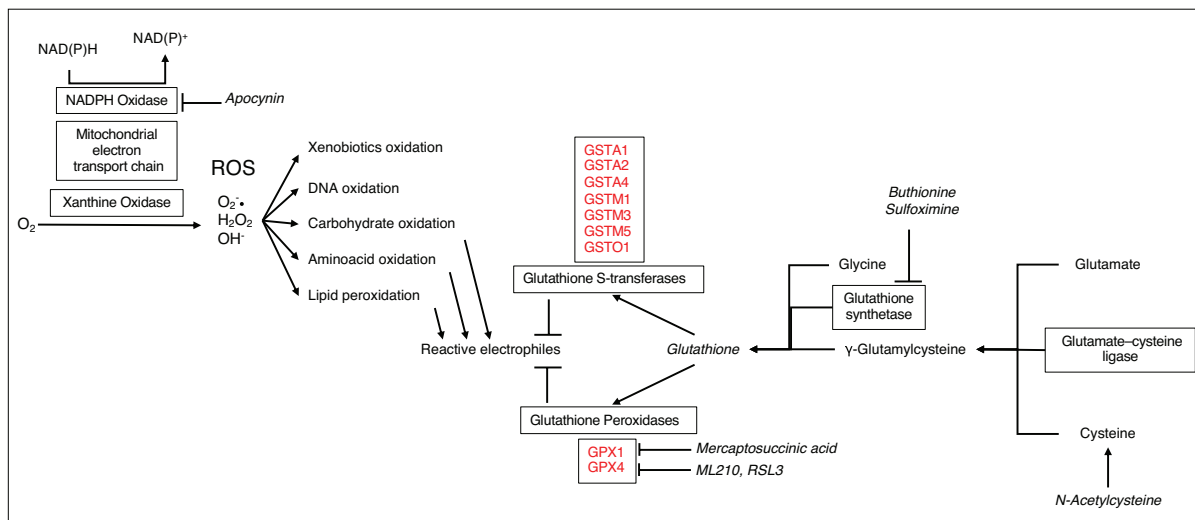
The data shown are representative of at least two independently performed experiments.



1300 **Fig. S7. sgRNA hits identified in purine metabolism pathway.**

Simplified representation of the purine metabolism pathway with significantly enriched (black) or depleted (red) sgRNA targets found in the genome-wide CRISPR/Cas9 screening. Drugs used to inhibit specific enzymes are *italicized*.

1305



1310

Fig. S8. ROS generation and scavenging.

Simplified representation of ROS production and scavenging pathways. Significantly depleted sgRNA targets found in the genome-wide CRISPR/Cas9 screening are annotated in red. Drugs and molecules used are *italicized*.

Kegg Pathway	KEGG ID	Freq	Genome-wide Ratio	Sample Ratio	p-value	FDR	q-value
Enriched Pathways							
Purine metabolism	hsa00230	5	0.060341821	0.263157895	0.004552013	0.022760067	ND
Mannose type O-glycan biosynthesis	hsa00515	2	0.007673526	0.105263158	0.009232434	0.023081085	ND
Depleted Pathways							
Glycolysis / Gluconeogenesis	hsa00010	19	0.023369376	0.049608355	0.002005235	0.033874609	0.01353054
Fatty acid degradation	hsa00071	12	0.014998256	0.031331593	0.014214189	0.067458054	0.02478919
Oxidative phosphorylation	hsa00190	25	0.041855598	0.065274151	0.020345478	0.067458054	0.02478919
Purine metabolism	hsa00230	36	0.060341821	0.093994778	0.006131678	0.047124933	0.01592725
Glutathione metabolism	hsa00480	14	0.019532612	0.036553525	0.020048346	0.067458054	0.02478919
Arachidonic acid metabolism	hsa00590	18	0.021276596	0.046997389	0.001710169	0.033874609	0.01353054
Metabolism of xenobiotics by cytochrome P450	hsa00980	17	0.025113359	0.044386423	0.018309849	0.067458054	0.02478919
Drug metabolism - cytochrome P450	hsa00982	18	0.023718172	0.046997389	0.005251754	0.047124933	0.01592725

Table S1. KEGG pathways related to statistically significant enriched or depleted

1315 **sgRNAs.**

Frequency refers to the number of unique significantly enriched/depleted gene-targets annotated in each pathway listed. The Genome-wide and Sample ratios are the proportion of genes within the human genome and unique significantly enriched/depleted gene-targets annotated within each pathway, respectively. Binomial enrichment testing provided p-values

1320 indicating most common pathways. Multiple testing correction was performed by computing the Benjamini-Hochberg false discovery rate (FDR) and q-value.

Recombinant DNA for sgRNA expression			
Target gene	gRNA sequence	Vector	Source
Human CRISPR Knockout Pooled Library (GeCKO v2)	-	-	Addgene Cat#1000000049
Cloning vector	-	lentiGuide-Puro	Addgene Cat#52963
ADA	CAGGCTTGATGGATCCGTCT	pLV-mCherry-U6 gRNA	VectorBuilder
ADA	TCACCGTACTGTCCACGCCG	lentiGuide-Puro	This paper
ADA	GCGGTACAGTCCGCACCTGC	lentiGuide-Puro	This paper
ADSSL1	TTCCAGGGGGCAACAACGC	lentiGuide-Puro	This paper
ADSSL1	GCTGATGATGTCCGGTCCG	lentiGuide-Puro	This paper
ADSSL1	ACATACCGAAGTCAATGTCG	lentiGuide-Puro	This paper
LACC1	CGTAGGTTGGCGAATGCTGC	lentiGuide-Puro	This paper
LACC1	TACCTGGGATCTCTCCGTT	lentiGuide-Puro	This paper
LACC1	TCAAGAAAATCTGCGTAGGT	lentiGuide-Puro	This paper
GLO1	GAACCGCAGCCCCCGTCCGG	pRP[gRNA]-EGFP:P2A:Puro-U6	VectorBuilder
GLO1	GTCCGGCGGCCTCACGGACG	pRP[gRNA]-EGFP:P2A:Puro-U6	VectorBuilder
TPI1	CGGCGAGGGCTTACCGGTGT	lentiGuide-Puro	This paper
TPI1	ACCGGTGTCGGCCGGCACCT	lentiGuide-Puro	This paper
TPI1	CGAAGTCGATATAGGCAGTA	lentiGuide-Puro	This paper
Scrambled sequence	GTGTAGTTCGACCATTCGTG	lentiGuide-Puro	This paper
Scrambled sequence	GTTACAGGATCACGTTACCGC	lentiGuide-Puro	This paper
Scrambled sequence	AAATGTGAGATCAGAGTAAT	lentiGuide-Puro	This paper
Recombinant DNA for gene overexpression			
Gene	Vector	Source	
LV-MR1A-β2m	Lentiviral overexpression vector	Addgene Cat#52962	
lentiCas9-Blast	Lentiviral overexpression vector	Addgene Cat#52962	
MR1 (NM_001531)	pET23d-MR1	Novagen, Cat#69748-3	
β2M (NM_004048.4)	pET23d-B2M	Novagen, Cat#69748-3	
hGLO1 (NM_006708.2)	pLV-EGFP:T2A:Puro-EF1A>hGLO1	VectorBuilder	
Oligonucleotides			
Name	Sequence	Source	
JScispr1	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNNNNNGCTT TATATATCTTGTGGAAAGGACGAAACACC	Microsynth	
JScispr3	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTGCCATTTGT CTCAAGATCTAGTTACGCCAAG	Microsynth	
Nextera Index Kit	-	Illumina Cat#15055290	

1325 **Table S2. Recombinant DNA used for knock-out generation and protein expression and oligonucleotides sequences.**

Data S1. (separate .xlsx file)

Significantly enriched sgRNAs classification according to KEGG pathway, Related to Figure 1 and Table S1.

1330 **Data S2. (separate .xlsx file)**

Significantly depleted sgRNAs classification according to KEGG pathway, Related to Figure 1 and Table S1.

8 Summary and Outlook

8 Summary and Outlook

8.1 Summary

8.1.1 Overview

In this PhD project, the toolbox of paramagnetic NMR was expanded step by step with the development of novel LCTs displaying significantly improved properties when compared to previous generations, i.e. enhanced induced anisotropy parameters, faster ligation kinetics, and reduction-stable linkages.¹⁻³ Subsequently, we developed an LCT equipped with an isothiocyanate activator for click chemistry on RNA molecules as well as an LCT, that is extremely immobilized on the protein's surface.⁴ Besides the design and realization of novel LCTs, a thorough study of the active site, mechanism and protein fold characteristics of the wild-type formylglycine-generating enzyme as well as two disease-related mutants identified in multiple sulfatase deficiency (MSD) patients was performed.⁵ Furthermore, the modification of purines and pyrimidines by reactive carbonyls, which accumulate in metabolically altered cancer cells, and the role of these nucleobase adducts as T cell antigens were investigated in collaboration with the research group of Prof. De Libero. Detailed summaries of the research projects included in this PhD thesis are given in the following.

8.1.2 Ln-DOTA-M8-(4*R*4*S*)-SSPy

In contrast to earlier generations of eight-fold methyl-substituted DOTA-derived LCTs,⁶ that can show two different conformational species in solution depending on the ionic radius of the lanthanide ion,⁷ the DOTA-M8-(4*R*4*S*)-SSPy LCT displays only one species over the entire series of lanthanides.¹ This is highly valuable for the acquisition of structural restraints, since the occurrence of only one conformational species significantly facilitates the assignment of the signals recorded in ¹H-¹⁵N HSQC experiments. The Tm-DOTA-M8-(4*R*4*S*)-SSPy LCT displays PCS of up to 5.5 ppm on ubiquitin S57C and selectively ¹⁵N leucine labeled human carbonic anhydrase II S50C constructs. ROESY experiments revealed that the lutetium and yttrium-variants of the DOTA-M8-(4*R*4*S*)-SSPy LCT exclusively adopt the square antiprismatic conformation. This was further investigated by DFT calculations of the applied paramagnetic lanthanide complexes (Tm, Dy and Tb) *in vacuo* and with implicit water solvent. The calculations strongly suggest that the Ln-DOTA-M8-(4*R*4*S*)-SSPy complexes adopt exclusively the square antiprismatic conformation (Figure 1), since the calculated energy differences between the two conformations amount to 7.0-50.5 kJ·mol⁻¹. This is in contrast to the (8*S*)-variant of the LCT, that shows a significant amount of both the SAP and TSAP conformer for the Dy- and Tb-complex (Dy: 73% SAP, Tb: 80% SAP).

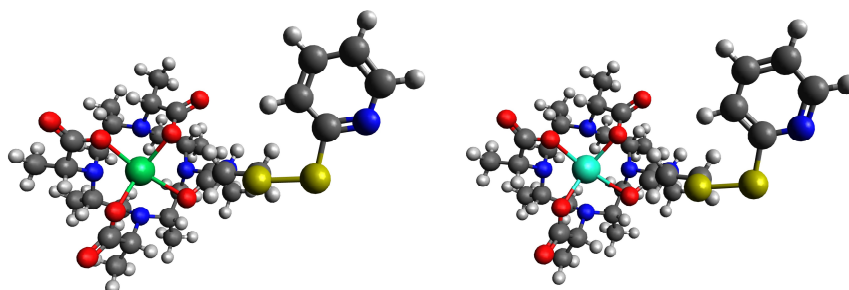


Figure 1: DFT structures of the SAP conformers of the thulium- and dysprosium-complex of DOTA-M8-(4*R*4*S*)-SSPy in vacuo (top view).

Our results for the (4*R*,4*S*)-variant were then confirmed by fitting of the signals in spectra of Yb-DOTA-M8-(4*R*4*S*)-SSPy to both the $\Lambda(\delta\delta\delta\delta)$ - and $\Delta(\delta\delta\delta\delta)$ -structure according to a procedure outlined in Gempf et al.⁸ Furthermore, the results are corroborated by the work of Ranganathan et al. concerning the ytterbium chelates of the complex with four lactic acid-derived side arms.⁹⁻¹⁰ The preference of the Ln-DOTA-M8-(4*R*4*S*)-SSPy complexes for the SAP conformation can be explained by the favorable alignment of the methyl groups within the chelate that leads to a minimized steric repulsion in the SAP conformation. The DOTA-M8-(4*R*4*S*)-SSPy LCT displays strong induced paramagnetic effects that are suitable for the study of biomacromolecules and allows for a convenient assignment of the measured signals. Furthermore, the LCT offers a less expensive synthetic route when compared with its predecessor with (8*S*)-configuration⁶, since the synthesis of the DOTA-M8-(4*R*4*S*)-SSPy LCT requires only readily available stereoisomers (*L*-Alaninol, *L*-Lactate) as starting materials for the synthesis.

8.1.3 Ln-P4M4-DOTA

Since the incorporation of large substituents into the ring scaffold of the LCT should increase the obtained paramagnetic effects by a decrease of motional averaging, we designed and synthesized an LCT incorporating isopropyl- instead of methyl-substituents in its scaffold.² The isopropyl-substituted version of cyclen was synthesized for the first time with a synthetic route that comprises 11 steps in total. The synthesis is able to deliver the macrocyclic scaffold in an overall yield of 6%. Subsequently, the installation of the sidearms and complexation of different lanthanide ions led to Ln-P4M4-DOTA. We showed by use of ROESY experiments that the Lu complex of P4M4-DOTA adopts exclusively a $\Lambda(\delta\delta\delta\delta)$ conformation in solution. Subsequently, DFT calculations yielded a stabilization of 32.6 kJ·mol⁻¹ for the Lu-variant and confirmed the exclusive presence of the $\Lambda(\delta\delta\delta\delta)$ conformation for the Tm- and Dy-complex. Interestingly, the highly rigidified LCT is able to induce large PCS of up to 6.5 ppm on ubiquitin S57C and selectively ¹⁵N labeled hCA II S166C constructs and leads to significantly enhanced PCS when compared to its predecessor Ln-DOTA-M8-(4*R*4*S*)-SSPy¹ (Figure 2).

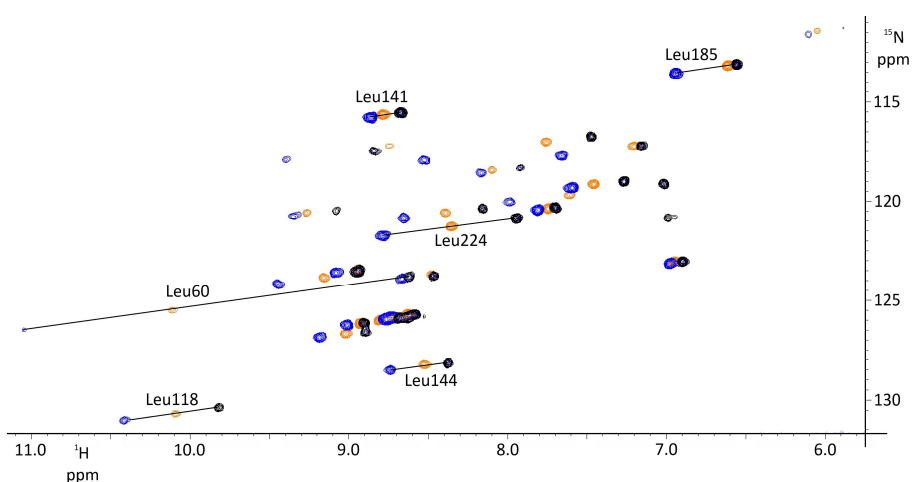


Figure 2: Overlay of ^1H - ^{15}N HSQC experiments of Dy-DOTA-M8-(4R,4S)-SSPy (orange), Dy-P4M4-DOTA (blue) and Lu-P4M4-DOTA (black) attached to selectively ^{15}N leucine labeled hCA II S166C. Visible from the spectra are the large obtained PCSs and the significant increase in PCS between the methyl- and isopropyl-substituted LCT loaded with dysprosium.

The increase in magnitude of observed PCSs originates in a decrease of the rotational averaging of paramagnetic effects on the surface of the protein. A significantly enlarged intrinsic anisotropy, translational averaging and different positioning of the donor atoms were excluded by 1D NMR experiments, a fitting routine to determine the metal center distribution by Suturina et al.¹¹⁻¹² and DFT calculations.

8.1.4 Ln-P4T-DOTA

In order to develop an LCT that provides a reduction-stable linkage to the biomacromolecule upon tagging, the isopropyl-substituted DOTA scaffold was equipped with a thiazole linker moiety (Ln-P4T-DOTA).³ The thiazole activator moiety leads to the formation of a reduction-stable thioether C-S bond upon conjugation to the cysteine residue of a protein and is, therefore, suitable for measurements under reductive conditions. Importantly, the combination of the isopropyl-substitution with a shorter linker when compared to Ln-P4M4-DOTA² leads to a significant reduction in motional averaging of the obtained paramagnetic effects. The novel LCT was benchmarked on ubiquitin S57C, ubiquitin K48C and selectively ^{15}N leucine labeled hCA II S166C and we found that it yields large PCS and very high anisotropy parameters ($\Delta\chi_{\text{ax}} = 54.0 \times 10^{-32} \text{ m}^3$ and $\Delta\chi_{\text{rh}} = 27.0 \times 10^{-32} \text{ m}^3$ for Dy on ubiquitin S57C at 298 K) (Figure 3).

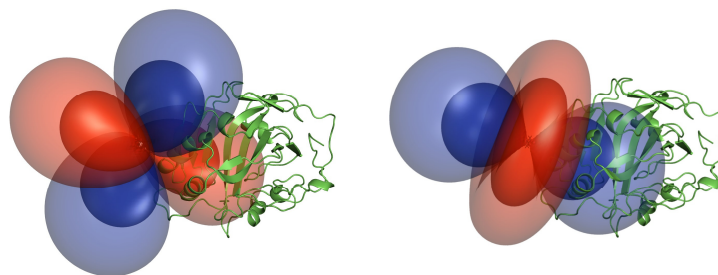


Figure 3: Obtained anisotropy parameters with Dy- (left) and Tm-P4T-DOTA (right) on a hCA II S166C construct visualized as PCS isosurfaces (0.5 ppm (outer layer), 2.0 ppm (inner layer)). Reproduced from Ref. ³ with permission from The Royal Society of Chemistry.³

The conformational bias towards a $\Lambda(\delta\delta\delta\delta)$ conformation was determined with DFT calculations and ROESY experiments. The LCT delivers large anisotropy parameters on all investigated protein constructs and would offer the possibility to be applied in in cell NMR measurements of proteins, due to the reduction-stable linkage formed between the LCT and the protein.

8.1.5 Ln-M7-Nitro

In order to further enhance the obtained anisotropy parameters and the size of detected, paramagnetic effects, a significant immobilization of the LCT on the protein's surface was required. The motional averaging of paramagnetic effects was already significantly decreased for the P4M4-DOTA² and P4T-DOTA LCT³, however, increased PCSs and RDCs are to be expected for a chelator that further restricts the motion of the LCT on the surface of the biomacromolecule. A research project of R. Vogel in our group and the estimate for the Tm-DOTA-M8-ITC LCT yielded intrinsic tensors for Tm³⁺ complexes of methyl-substituted DOTA LCTs with a magnitude around $\Delta\chi_{ax} = 100 \times 10^{-32} \text{ m}^3$. Since the Tm-P4T-DOTA LCT shows a maximal axial anisotropy parameter of $\Delta\chi_{ax} = 44.3 \times 10^{-32} \text{ m}^3$ on a hCA II S166C construct, the full potential of methyl-substituted DOTA-derived LCTs remains therefore to be unlocked by an LCT that shows a strongly suppressed mobility on the surface of the protein. In order to design such a positional invariant LCT, a nitro-substituted pyridine activator with the conjugation site for the protein's cysteine in *ortho* position was used as linker (Ln-M7-Nitro).⁴ The conjugation of the cysteine in the *ortho* position provides, due to the increased steric clashes of the LCT with the protein, a strong suppression of the rotational averaging when compared to a ligation in the *para* position (Figure 4).

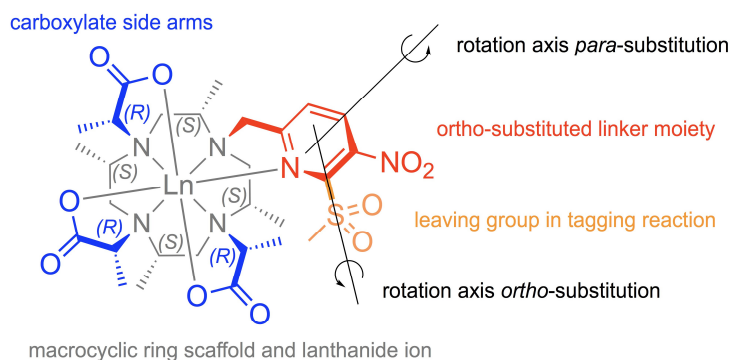


Figure 4: Chemical structure of the Ln-M7-Nitro LCT in $\Lambda(\delta\delta\delta\delta)$ conformation and indication of the relevant structural parts and features. Reproduced from Ref. ⁴ with permission from The Royal Society of Chemistry.⁴

The nitro group stabilizes the Meisenheimer intermediate during the tagging reaction and delivers thereby an LCT that displays highly favorable ligation kinetics, i.e. tagging of the cysteine residue of the investigated protein constructs in an extent greater than 95% over 30 minutes under close to physiological conditions (highly diluted (90 μ M), aqueous buffer, pH 7.0). The novel, rationally designed LCT induces unprecedented anisotropy parameters on both ubiquitin S57C and hCA II S50C constructs ($\Delta\chi_{ax} = 95.9$ and $\Delta\chi_{rh} = 13.1$ for Dy on hCA II S50C at 298 K). This can be attributed to the extremely short linkage between the LCT and the protein that ensures a strong immobilization of the LCT on the protein's surface. Due to the large induced anisotropy parameters, reduction-stable linkage and favorable ligation kinetics, the LCT provides a solution for the measurement of structural restraints over increased distances.

8.1.6 Ln-DOTA-M8-ITC

Due to the high interest in the structural characterization of RNA targets for future drug design and the current prominent role of NMR thereby, it constitutes a highly tempting target to develop an LCT suitable for acquisition of long-range restraints on RNA molecules and their complexes.¹³⁻¹⁵ To achieve this goal, we proposed and synthesized a novel RNA LCT with an isothiocyanate activator moiety (Ln-DOTA-M8-ITC). The reactive isothiocyanate moiety is conjugated to an aminomethyl-modified uridine base on the RNA strand and results in a stable, covalent thiourea linkage between the LCT and the RNA molecule. In order to obtain an insight into the suitability of the Ln-DOTA-M8-ITC LCT for acquisition of paramagnetic restraints, the LCT was benchmarked on a 36-nt strand. PCSs were acquired with ^1H - ^{15}N and ^1H - ^{13}C HMQC/HSQC as well as HNN COSY experiments. By subsequent fitting of the obtained PCSs, we showed that the magnitude and the Q-factor of the induced anisotropy parameters with the Tm-DOTA-M8-ITC LCT ($\Delta\chi_{ax} = 21.5$ and $\Delta\chi_{rh} = 8.7$, $Q = 0.072$) are comparable to those

obtained with Tm-DOTA-M8-(4*R*,5)-SSPy¹ on a ubiquitin S57C construct ($\Delta\chi_{ax} = 19.6$ and $\Delta\chi_{rh} = 3.0$, $Q = 0.062$) (Figure 5).

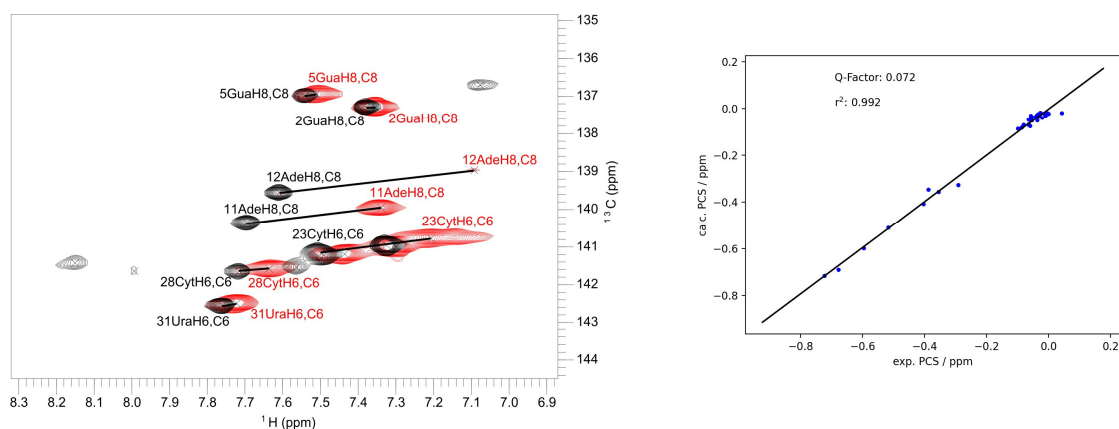


Figure 5: Selected PCSs induced by the Tm-DOTA-M8-ITC LCT measured by ¹H-¹³C TROSY HSQC experiments on the 36-nt RNA strand (left) and correlation plot of experimental and back-calculated PCSs (right).

The obtained results demonstrate that the PCS methodology was successfully transferred from the protein to the RNA world and we are looking forward to numerous emerging future applications for the investigation of structure, dynamics and ligand-binding of RNA molecules.

8.1.7 Investigation of the structure, active site and dynamics of formylglycine-generating enzyme

Due to the special fold and catalytic mechanism of the FGE protein, the enzyme constitutes a highly interesting target for NMR studies and we investigated structure, dynamics and mechanism of the FGE protein in this PhD project. The CSPs of the amide N-H spin pairs in Ag(I) and Cu(I) containing FGE variants as well as of different mutants were extracted from ¹H-¹⁵N HSQC experiments. Together with a novel FGE structure acquired by X-ray crystallography (resolution = 1.04 Å), we were able to obtain an insight into the unusual active site of the FGE protein, which consists of a copper ion coordinated by two cysteine residues in a bislinear motif.⁵ Upon binding of a 17-residue analog for the natural substrate of the FGE protein, the active site rearranges to a trigonal planar Cu(I) complex with a juxtaposed O₂-binding site. The NMR assignment of both the Ag(I) and Cu(I)-containing uniformly ²H, ¹³C, ¹⁵N labeled FGE construct offered the possibility to investigate the residue-specific flexibility and stability by means of T₁ as well as T₂ relaxation, heteronuclear Overhauser effect and amide exchange experiments (Figure 6).

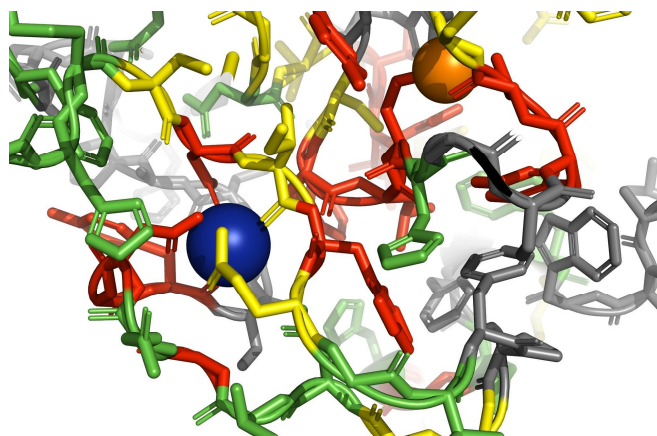


Figure 6: Visualization of the rigidification of the outer Ca(II) binding site (lower left, Ca(II) ion in dark blue) and the connection and rigidification of the active site (upper right, Ag(I) ion in orange) obtained via measurements of T_1 relaxation in uniformly ^2H , ^{13}C , ^{15}N -labeled wild-type FGE containing Ag(I). Green residues: $T_1 > 830$ ms, yellow residues: 830 ms $> T_1 > 780$ ms, red residues: $T_1 < 780$ ms, grey residues: not assigned or analysis impossible due to signal-to-noise ratio.

Subsequently, with the assignment in hand, we investigated two disease-related mutants of the FGE protein and, therefore, the molecular basis of MSD for the two mutations under investigation. Most importantly, we showed that the enzyme is destabilized by the loss of one or two Ca(II) ions and displays a strikingly lower Cu(I)-affinity for the disease-related N188S and N188I variants. The conducted studies of the FGE protein by use of solution NMR shed light on the structural origin of MSD and the properties of this highly special and exciting copper-containing enzyme.

8.1.8 T cell activation by nucleobase adduct-containing metabolites

Antigens in human cells can be bound and presented on the surface by proteins as e.g. the non-polymorphic MHC-class-I-related molecule MR1.¹⁶ Upon binding of the antigen, structural changes within the protein occur that give rise to binding by T cells.¹⁷ We investigated, if malondialdehyde adducts of nucleotides, which occur in cancer cells due to the altered metabolism, are able to trigger a response by the immune system. Four different adducts of malondialdehyde with guanine, adenine and cytosine were synthesized and characterized by NMR and HRMS in this thesis project (Figure 7) and subsequently tested for both the display on the surface of cells by MR1 and recognition by T cells.

in solution. Subsequently, the Ln-M7-Nitro LCT was developed and due to the extreme immobilization of the LCT on the surface of the protein, exceedingly large anisotropy parameters were obtained and the LCT should deliver structural restraints over distances of up to 200 Å in favorable cases. The significantly shortened linker, as well as the possibility to record NMR experiments under reducing conditions, render the Ln-M7-Nitro LCT as an ideal LCT for acquiring structural restraints over increased distances. The development of LCTs for proteins has reached a very high level nowadays. However, there are still interesting developments that can be made. The limited supply of fourfold methyl-substituted cyclen renders LCTs that are based on an alternative scaffold, which is either commercially available or more simple to obtain, very useful. If one could create an LCT, that is based on an easily accessible scaffold, with anisotropy parameters in the range of methyl-substituted, DOTA-derived LCTs, the application of LCTs in biomolecular NMR would be significantly pushed forward. Besides the development of LCTs based on simple scaffolds, a single-armed LCT that leads to predictable anisotropy tensors would be highly favored by the community. The independence from the local environment of the protein, as e.g. demonstrated for the two-armed CLaNP-5 LCT, leads to predictable anisotropy tensors.¹⁸ Thereby, a large part of the experimental work, i.e. the tagging and evaluation of PCSs for the fitting of the anisotropy parameters, can potentially be omitted. Depending on the goal of the study and accuracy of the needed structural restraints, this can lead therefore to the direct start of the measurement of PCS for the structure elucidation, e.g. the localization of a binding partner, and thus with the exploration of biologically relevant questions.¹⁹

For the first time, large PCS induced by an LCT were detected on an RNA target. This was achieved by a rationally designed isothiocyanate moiety on the LCT that can be conjugated to an aminomethyl-uridine modified nucleobase in the RNA strand. The ligation reaction proceeds under ambient temperature to completion within 5 h. The novel LCT was benchmarked on a 36-nt RNA construct. Thereby, we found that the Tm-DOTA-M8-ITC LCT induces large PCSs and anisotropy parameters comparable to Tm-DOTA-M8-(4*R*4*S*)-SSPy (on ubiquitin S57C). These results demonstrate that the toolbox of paramagnetic NMR was successfully transferred from proteins to RNA. Currently, a 65-nt RNA strand and the position of its ligand, *S*-adenosyl methionine, are investigated with the Ln-DOTA-M8-ITC LCT. Furthermore, MD simulations are planned to be performed in order to get an insight into the dynamics of the 36-nt RNA-LCT conjugate. Based on 1) the estimated, intrinsic anisotropy parameters on the free tag, 2) the averaged, effective anisotropy parameters on the RNA-LCT conjugate and 3) the motional data from MD simulations, it is expected to obtain a realistic picture of the dynamics present in the 36-nt RNA-LCT system. In the future, many novel methyl-substituted, DOTA-derived LCTs can be designed that take advantage of the developed isothiocyanate activator moiety. Most promising would be LCTs with a shorter linker than Ln-DOTA-M8-ITC, e.g. an analog that

contains a pyridine-isothiocyanate activator moiety in its structure. The aromatic ring equipped with the isothiocyanate activator is not fixed to the scaffold in Ln-DOTA-M8-ITC, however, the pyridine ring in an improved version of the LCT (Ln-DOTA-M7Py-ITC) could with its nitrogen atom coordinate to the metal and thereby fixedly attach the linker moiety to the metal center (Potential synthetic route given in Figure 8. Structure of the LCT on the bottom, right). This additional rigidification should lead to significantly enlarged paramagnetic effects and thus to an increased size of the observed paramagnetic effects.

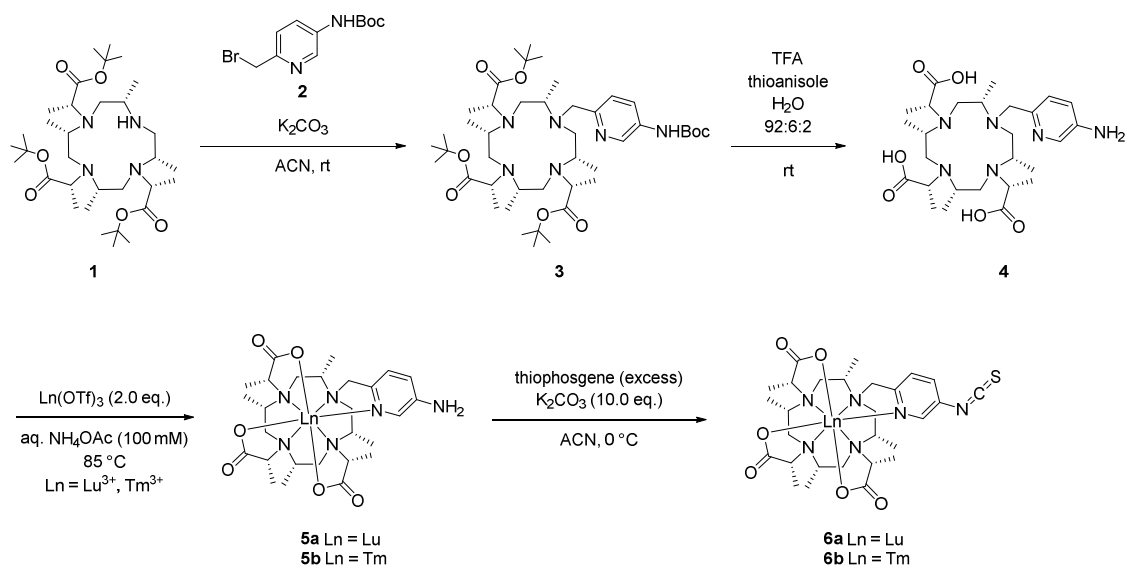


Figure 8: Synthetic route to an isothiocyanate equipped, pyridine-derived LCT (Ln-DOTA-M7Py-ITC) for tagging of amino groups incorporated in RNA bases.

In addition to the design and synthesis of more advanced generations of Ln-DOTA-M8-ITC, a large range of applications for the new LCT could be implemented. Challenging applications in the investigation of RNA-RNA, RNA-protein and additional RNA-ligand complexes and elucidation of the structure of RNA targets would be highly suitable.

8.2.2 Investigation of the structure, active site and dynamics of formylglycine-generating enzyme

In the FGE project, structural data from X-ray crystallography and NMR was analyzed. Based on the FGE crystal structure with a resolution of 1.04 Å and ^1H - ^{15}N HSQC experiments measured by NMR spectroscopy, we monitored and characterized the interaction of the FGE protein with a 17-residue substrate analog. We showed thereby that the Cu(I) coordination by two cysteines in bislinear fashion changes to a trigonal planar Cu(I) complex upon binding of the substrate analog. Most importantly, an acidic pocket for the incoming oxygen during catalysis is formed. However, there is no direct binding site for the incoming oxygen on the metal center formed. Thereby, an outer sphere electron transfer from Cu(I) to the oxygen is

performed that delivers the superoxide species required for catalysis. Subsequently, two disease-related mutants identified in MSD patients were studied by solution NMR (N188S and N188I). This required an assignment of the functional state of the FGE species with one and two incorporated Ca(II) ions, respectively. Thereby, we showed that the N188S mutant contains only the outer Ca(II) ion and the N188I mutant lost both Ca(II). This is in contrast to the native FGE protein that contains two Ca(II) ions and leads to a destabilization of the mutants. Interestingly, the high proline content (~10%) present in the protein could have severe implications for both the dynamics of the fold as well as the entropic stabilization of the folded protein when compared to the unfolded state. An increase in proline content is associated with a stability gain of the folded protein with respect to the unfolded state. This originates in a lower entropy of the unfolded protein chain incorporating a high proline content when compared to an amino acid chain with less proline content. Furthermore, a highly interesting topic would be the exploration of the role of protein dynamics in the catalysis of the FGE protein, in general and with respect to different proline conformations.

8.2.3 T cell activation by nucleobase adduct-containing metabolites

The metabolic state of cancer cells leads to an increase of reactive oxygen species and, therefore, to an increased presence of nucleobase adducts within the cells. In our study, we synthesized four nucleobase-malondialdehyde adducts and investigated their presentation to T cells by the MHC-class-I-related molecule MR1. Thereby, we showed that the nucleobase-adducts are both displayed on the surface and recognized by the immune system. The fact that the metabolic changes that lead to generation of such adducts are specifically found in cancer cells implies that a novel therapy based on training of the immune system could be implemented. In the future, studies involving additional adducts of molecules with reactive metabolites present in cancer cells could lead to an extension of the knowledge about cancer-specific T cell antigens. Furthermore, more detailed studies about the interface of MR1 and T cells could be performed by use of NMR spectroscopy in order to obtain insights into the interactions and opportunities for novel therapeutics imitating the MR1-ligand complex.

References

1. D. Joss, R. M. Walliser, K. Zimmermann, D. Häussinger, *J. Biomol. NMR*, **2018**, 29-38.
2. D. Joss, M.-S. Bertrams, D. Häussinger, *Chem. Eur. J.*, **2019**, *25*, 11910-11917.
3. D. Joss, D. Häussinger, *Chem. Commun.*, **2019**, *55*, 10543-10546.
4. D. Joss, F. Winter, D. Häussinger, *Chem. Commun.*, **2020**, *56*, 12861-12864.
5. D. A. Miarzlou, F. Leisinger, D. Joss, D. Häussinger, F. P. Seebeck, *Chem. Sci.*, **2019**, *10*, 7049-7058.
6. D. Häussinger, J.-R. Huang, S. Grzesiek, *J. Am. Chem. Soc.*, **2009**, *131*, 14761-14767.
7. A. C. L. Opina, M. Strickland, Y.-S. Lee, N. Tjandra, R. A. Byrd, R. E. Swenson, O. Vasalatiy, *Dalton Trans.*, **2016**, *45*, 4673-4687.
8. K. L. Gempf, S. J. Butler, A. M. Funk, D. Parker, *Chem. Commun.*, **2013**, *49*, 9104-9106.
9. R. S. Ranganathan, R. K. Pillai, N. Raju, H. Fan, H. Nguyen, M. F. Tweedle, J. F. Desreux, V. Jacques, *Inorg. Chem.*, **2002**, *41*, 6846-6855.
10. R. S. Ranganathan, N. Raju, H. Fan, X. Zhang, M. F. Tweedle, J. F. Desreux, V. Jacques, *Inorg. Chem.*, **2002**, *41*, 6856-6866.
11. E. A. Sutura, D. Häussinger, K. Zimmermann, L. Garbuio, M. Yulikov, G. Jeschke, I. Kuprov, *Chem. Sci.*, **2017**, *8*, 2751-2757.
12. E. A. Sutura, I. Kuprov, *Phys. Chem. Chem. Phys.*, **2016**, *18*, 26412-26422.
13. D. Joss, D. Häussinger, *Prog. Nucl. Magn. Reson. Spectrosc.*, **2019**, *114-115*, 284-312.
14. B. Fürtig, C. Richter, J. Wöhnert, H. Schwalbe, *ChemBioChem*, **2003**, *4*, 936-962.
15. D. Joss, R. Vogel, D. Häussinger, *Ref. Mod. in Chem, Mol. Sci. and Chem. Eng.*, **2020**, doi:10.1016/B978-0-12-409547-2.14848-6.
16. M. Lepore, L. Mori, G. De Libero, *Front. Immunol.*, **2018**, *9*, 1365.
17. K. Natarajan, J. Jiang, N. A. May, M. G. Mage, L. F. Boyd, A. C. McShan, N. G. Sgourakis, A. Bax, D. H. Margulies, *Front. Immunol.*, **2018**, *9*, 1657.
18. P. H. J. Keizers, A. Saragliadis, Y. Hiruma, M. Overhand, M. Ubbink, *J. Am. Chem. Soc.*, **2008**, *130*, 14802-14812.
19. J.-Y. Guan, P. H. J. Keizers, W.-M. Liu, F. Löhr, S. P. Skinner, E. A. Heeneman, H. Schwalbe, M. Ubbink, G. Siegal, *J. Am. Chem. Soc.*, **2013**, *135*, 5859-5868.