Development and application of an economic approach to isotope labeling in higher eukaryotes for NMR studies

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Christian Opitz

aus Rümmingen, Deutschland

Basel, 2018

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät
auf Antrag von

Prof. Dr. Stephan Grzesiek
Prof. Dr. Timm Maier

Basel, den 19.04.2016

Prof. Dr. Jörg Schibler
(Dekan)

Summary

Cutting-edge structural studies by crystallography are presently aimed at very complex targets including membrane, multi-domain and secreted proteins of human origin. Such proteins can often only be expressed in higher eukaryotic systems, such as insect or mammalian cell lines due to the requirement of the complex eukaryotic protein expression and post-translational modification machinery. For NMR studies, the need of isotope labeling poses extreme challenges in such eukaryotic hosts, since the growth media need to contain isotope-labeled amino acids. The costs of commercially available isotope-labeled amino acids are very high and so far even prohibitive in the case of deuteration.

This thesis describes the development of an economic approach to isotope labeling in higher eukaryotes, where the necessary labeled amino acids are provided to the growth media from autolyzed yeast grown on inexpensive ^{13}C glucose, ^{15}N ammonium salt and $^{2}H_2O$. The method was used in the baculovirus-infected insect cell system to produce isotope-labeled proteins relevant to human health and their study by NMR.

Chapter 1 provides an introduction to recombinant expression in insect cell lines and the use of the baculovirus expression vector system. Furthermore, the amino acid metabolism in insect cells and its relevance for isotope labeling is discussed. Frequently used approaches to isotope labeling and their applications are outlined.

Chapter 2 describes in detail the production of isotope-labeled yeast extracts and their supplementation to insect cell growth media for labeled protein expression. Parts of this chapter have been published (Opitz *et al*., An economic approach to efficient isotope labeling in insect cells using homemade ^{15}N-, ^{13}C- and ^{2}H-labeled yeast extracts. *J. Biomol. NMR*, 2015). The method achieves uniform protein ^{15}N and ^{13}C labeling of up to 90 % and deuteration levels of >60 %, which are the highest reported incorporation levels based on cellular extracts so far.

Chapter 3 describes the design of a thermostabilized mutant of the turkey β_1-adrenergic receptor (TS-β_1AR) for NMR solution studies. This mutant was expressed with selective ^{15}N-valine labeling in insect cells and used for an extensive NMR characterization of ligand interactions. It was possible to follow ligand-induced backbone motions at many key residues in the receptor. The latter results have been published (Isogai, Deupi, Opitz *et al*., Backbone NMR reveals allosteric signal transduction networks in the β_1-adrenergic receptor. *Nature*, 2016). The final section of the chapter shows the expression of uniformly ^{2}H-^{15}N labeled TS-β_1AR, which has been used to record high-quality ^{15}N-edited NOESY spectra and obtain a number of further backbone resonance assignments.

Chapter 4 describes the expression of the human lipopolysaccharide-binding protein (hLBP) for NMR studies in baculovirus-infected insect cells by secretion of the protein to the culture medium. A labeling scheme was explored based on the combination of isotope-labeled autolyzed yeast and commercial extracts of algal cells. The combined approach made it possible to produce milligram quantities of uniformly ^{15}N-labeled hLBP and to acquire two-dimensional ^{1}H-^{15}N-TROSY spectra.

Chapter 5 describes the enzymatic synthesis of $^{15}N_2$-L-glutamine from inexpensive 2-oxoglutaric acid and ^{15}N ammonium salt. Since supplementation of L-glutamine is required for efficient expression, the method further reduces the costs of the labeling approach based on autolyzed yeast and enables economic production of selectively deuterated L-glutamine.

For additional cost reduction, also the recombinant expression of glutamate-ammonia ligase was established.

Chapter 6 describes the use of autolyzed yeast as an economical source of labeled amino acids in the cell-free expression system. An initial protocol for high-yield expression was established and its optimization is discussed. Isotope labeling by ^{15}N and ^{2}H is demonstrated for the green fluorescent protein GFP produced in the cell-free expression system based on isotope-labeled yeast.

Chapter 7 describes a proteomics approach to study short-term adaptation of *E. coli* to growth on $^{2}H_7$-glucose and $^{2}H_2O$. Characterization of the cellular adaptation process has implications for the production of deuterated proteins. The results show that deuterated media induce differential expression of proteins involved in the response to osmotic and temperature-induced stress. The data further indicate downregulation of proteins involved in ribosome biogenesis and differential activation of chaperones targeting nucleic acids as well as proteins. Simultaneously, an alteration of the energy metabolism was observed in the perdeuterated environment, in particular upregulation of anaplerotic reactions in the TCA cycle.

Chapter 8 explores the use of isotope-labeled yeast as an economic source of amino acids for labeling in mammalian cells. The initial results indicate that, similar to the protocol applied in insect cells, inexpensive dropout media supplemented by labeled yeast extracts enable good expression of the uniformly labeled angiotensin II receptor type 1a for NMR studies.

Results from this thesis have been published:

1. C. Opitz, S. Isogai, S. Grzesiek, An economic approach to efficient isotope labeling in insect cells using homemade (15)N-, (13)C- and (2)H-labeled yeast extracts. *J. Biomol. NMR*. **62** (3), 373-285 (2015).

2. S. Isogai, X. Deupi, C. Opitz, F. M. Heydenreich, C.-J. Tsai, F. Brueckner, G. F. X. Schertler, D. B. Veprintsev, S. Grzesiek, Backbone NMR reveals allosteric signal transduction networks in the β1-adrenergic receptor. *Nature*. **530**, 237–241 (2016).

The following further publications are planned:

1. C. Opitz, S. Isogai and S. Grzesiek, Tracking ligand-induced conformational changes in uniformly labeled β_1-adrenergic receptor by NMR (Chapter 3)

2. C. Opitz, S. Reckel, H. Aragao, S. Hiller and S. Grzesiek, An economic protocol for cell-free protein synthesis based on isotope-labeled yeast extracts (Chapter 6),

3. C. Opitz, E. Ahrne, A. Schmidt and S. Grzesiek, Short-term adaptation of E. coli to growth in a perdeuterated environment probed by proteome-wide analysis (Chapter 7),

4. C. Opitz, C. L. Piscitelli, G. F. X. Schertler and S. Grzesiek, Uniform isotope labeling of the human neurotensin receptor 1 expressed in HEK293 cells (Chapter 8).

Acknowledgments

I would like to thank everyone who by his or her willing support and help, however small, made this thesis possible. I had the great pleasure to work with very talented and remarkable individuals throughout the last four years. First and foremost, I would like to express my sincere gratitude to my supervisor Prof. Dr. Stephan Grzesiek for giving me the opportunity and liberty to conduct my thesis in a highly stimulating environment. His guidance, eagerness to share his knowledge and his enduring support throughout this work is highly appreciated. I recall that it was actually Stephan, who urged me to pursue a PhD and continued to inspire me with an unparalleled passion for science. My experience in his research group has been nothing short of amazing. I also would like to thank all the past and present members of his group for challenging and enriching my ideas.

I must offer my profoundest gratitude to my thesis committee members Prof. Dr. Timm Maier and Prof. Dr. Kurt Ballmer-Hofer for being major advisors and guidance. This includes all insightful comments on my work and the fair evaluation of my progress. In particular, I am grateful to Timm for kindly accepting to co-referee my thesis.

This thesis is the result of a combined effort and I have greatly benefited from the curious and helpful nature of several people. I highly appreciated the hard work by Jethro Hemmann, which provided further insights into isotope labeling in insect cells. You have been very well organized and thinking alike. I also praise the enormous amount of help and skillful hands of Ines Hertel-Hering, who has taken great pleasure in supporting me with all my challenging tasks in the wet lab that seemed so trivial to her. The incessant work by Dr. Shin Isogai on the β_1-adgrenergic receptor is extensively acknowledged and has been instrumental to obtain some beautiful results. I am also grateful for all our discussions and your support.

Further, I wish to thank Leonildo Delgado for performing the challenging tasks of expressing and purifying the lipopolysaccharide binding protein (LBP). I must offer my gratitude to Marco Rogowski for his significant contribution regarding the enzymatic synthesis of glutamine. Marco's efforts also guaranteed the smooth running of several instruments in our laboratory. Dr. Timothy Sharpe is gratefully acknowledged for his assistance with the light-scattering measurements on the recombinant glutamine-ammonia ligase. I am obliged to Dr. Sina Reckel for your motivation and the boundless help with the cell-free expression system. I very much appreciated the good company and your enthusiasm. I would also like to acknowledge the proteomics core facility in the persons of Drs. Erik Lennart Ahrné and Alexander Schmidt for making it possible to study the effect of deuterium on *E. coli*. Many thanks for the patience and the effort you put in all the exotic tasks I have been imposing on you. I am also very grateful for the help by Dr. Paul Jenö and Suzette Moes, who have not been hesitating to jump the sample queue for me. I have been very privileged to get to know Dr. Chayne Piscitelli with whom I have been and hope to continue working on the extension of the labeling protocol to mammalian cells. You are certainly a meticulously working scientist.

Furthermore, I would like to single out PD Dr. Daniel Häussinger, who has been a great source of expertise and a jack-of-all-trades in organic chemistry. Thank you for sharing your resources including a very limited amount of time, a spectrometer and vital parts of your laboratory. Jan Welker and Dr. Adam Mazur, thank you for providing neat solutions for often intricate IT problems. Dr. Xavier Deupi deserves a special mention for his valuable hint at the making of GPCR snake plots. Drs. Hans-Jürgen Sass and Judith Maria Habazettl deserve my

special thanks for passing on their NMR knowledge - who said you couldn't teach a young dog some old tricks? And finally, I must acknowledge Dr. Lydia Nisius for laying a fruitful foundation for my thesis work before I even started.

I am particularly grateful to Drs. Antoine Gautier and Hoi Tik Alvin Leung for proofreading parts of this thesis and their helpful comments. My sincere thanks go to Alvin for reallocating his knowledge of Python to me. I wish the two of you all the best for your future endeavors.

Lastly, I am deeply grateful for all the support from my family and friends. Ever since, they have provided a home and have believed in my pursuit, however stubborn. I am forever indebted to my parents, who shaped my mind and invested so much for my education. Very special thanks to Drs. Lukasz Stanislaw Skora and Petranka Krumova for their continuous encouragement and stimulating discussions on the use of yeast outside of a bioreactor. I was blessed to meet the two of you and hope for many more years ahead. Most of all, I owe my utmost gratitude to my wife Saskia Braune for her unwavering love and unconditional support throughout this journey – in all that exists none has your beauty. I am fortunate to share my life with Bruno and Mascha, who have welcomed me in their hearts and minds. Our son Edgar for gifting his parents with the greatest present of all: a future. Thank you.

Contents

Abbreviations and symbols

Å	0.1 nm
aa	amino acid
AATS	amino acid-type selective
AE	algal extract
ATP	adenosine triphosphate
BEVS	baculovirus expression vector system
cAbl	Abelson tyrosine kinase domain
CFE	cell-free expression
CHO	Chinese hamster ovary cells
CHS	cholesteryl hemisuccinate
Da	dalton, g mol^{-1}
DDM	*n*-dodecyl β-D-maltopyranoside
DM	*n*-decyl β-D-maltopyranoside
DMEM	Dulbecco's Modified Eagle's Medium
DNA	deoxyribonucleic acid
E. coli	Escherichia coli
EDTA	ethylenediaminetetraacetic acid
FCS	fetal calf serum
FSEC	fluorescence-detection size-exclusion chromatrography
GFP	green fluorescent protein
GPCR	G protein coupled receptor
HEK293	human embryonic kidney 293 cells
HPLC	high performance liquid chromatography
HSQC	heteronuclear single quantum coherence
ICL	insect cell lines
INEPT	insensitive nuclei enhancement by polarization transfer
K_M	Michaelis constant
LB	lysogeny broth
LBP	lipopolysaccharide binding protein
LC	liquid chromatography
LC/MS	liquid chromatography-mass spectrometry
LPS	lipopolysaccharide
MALS	multi-angle light scattering
MCL	mammalian cell lines
MS	mass spectrometry
MS/MS	tandem mass spectrometry
MWCO	molecular weight cut-off
NAD	nicotinamide adenine dinucleotide
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser enhancement spectroscopy
ORD	optical rotatory dispersion
PAGE	polyacrylamide gel electrophoresis

PEM	protein expression medium
PDB	RCSB Protein Data Bank (www.rcsb.org/pdb)
PITC	phenyl isothiocyanate
ppb	parts per billion
ppm	parts per million
RNA	ribonucleic acid
SDC	sodium deoxycholate
SDS	sodium dodecyl sulfate
SEC	size-exclusion chromatography
TCA	tricarboxylic acid cycle
TM	transmembrane
TROSY	transverse relaxation optimized spectroscopy
V_{max}	maximal velocity
YE	yeast extract

1 Introduction

1.1 Scope of this thesis

Production of isotope-labeled proteins for nuclear magnetic resonance studies

Ambitious contemporary structural and functional studies by nuclear magnetic resonance (NMR) are aimed at complex targets including membrane, multi-domain and secreted proteins of human origin. The production of these proteins in functional isotope-labeled form is a major bottleneck. *Escherichia coli* (*E. coli*) remains the most widely used expression system offering a number of advantages, namely inexpensive culture media, rapid growth, ease of cloning, and most importantly, a wide variety of isotope labeling schemes. These labeling schemes have been developed based on synthetically defined minimal media, which are often inexpensive and enable very high incorporation levels. Key to this development is the metabolic flexibility of *E. coli*. In fact, growth and expression can be established on simple carbon and nitrogen sources such as ^{13}C-glucose and ^{15}N-ammonium salts. Further, *E. coli* is amenable to growth in a perdeuterated environment and provides efficient deuteration (>90 %) (*1*, *2*) to improve the spectral properties of larger proteins (>20 kDa).

However, correct folding and activity of human proteins often requires various post-translational modifications, which are highly limited or even absent in *E. coli*. Especially, erroneous formation of disulfide bonds poses a major challenge and frequently results in aggregation and the loss of biological function (*3*). Furthermore, efficient routing of transmembrane and secretion of glycoproteins is often not possible. Great effort has been made to improve folding and expression of eukaryotic proteins in *E. coli* including alteration of the cellular redox environment (*4*), the use of chaperones (*5*), and glycol-engineering (*3*, *6*). Although these approaches yield impressive results for some targets (*7*), they are not applicable to all proteins.

In contrast, insect cell lines (ICL) have been widely used for the recombinant expression of several thousands challenging higher eukaryotic proteins for crystallographic studies (*8*). Key to this success is the cellular environment of the insect cell, which comprises a comprehensive folding machinery and provides post-translational modifications such as homogenous protein glycosylation patterns. Most notably, baculovirus-infected insect cells have played a pivotal role in crystallographic studies of G protein coupled receptors (GPCRs) (*9*) and have been applied for the determination of 84 GPCR structures out of 115 known to date (*10*). Using the baculovirus expression vector system (BEVS), high amounts of protein can be obtained by using strong viral promoters that encode for proteins, which are not essential for proliferation of the baculovirus and thus can drive recombinant expression of heterologous genes. Moreover, suspension cultures of ICL display excellent scalability and compare well to flask-based expression in *E. coli*. Large-scale production of recombinant proteins is commonly achieved by cultivation in disposable conventional bioreactors (*11*) or the efficient WAVE Bioreactor™ System (*12*).

However, production of isotope-labeled proteins for NMR studies poses a major challenge in insect cells and in fact, in all higher eukaryotic expression systems. The limited metabolic capabilities of insect cells do not allow uniform isotope labeling from growth on simple carbon and nitrogen sources, which are successfully used for low-cost recombinant expression in *E. coli* and lower eukaryotes such as yeast. Consequently, the concept of a defined minimal medium is not available for insect cell cultures and uniform labeling has so far been exclusively relying on very expensive commercial media (*13-15*).

Most importantly, perdeuteration of non-exchangeable proton sites is key to NMR studies of complex proteins (*1*, *2*). This requires growth of insect cells on culture media supplemented with deuterated amino acids. However, as such media contain several hundred milligrams of individual amino acids, the costs are prohibitively high. Indeed, a commercial medium for uniform deuteration in insect cells is currently not available.

Aims of this work

Although commercial media for ^{15}N and ^{13}C, ^{15}N labeling in insect cells have been available for nearly a decade, routine use has not yet become widespread. This is in part due to the very high costs associated with this approach and more importantly, the lack of deuteration, which is an essential tool to study complex proteins by NMR. An alternative approach has gained increased momentum (*16*), which is based on the use of isotope-labeled yeast extracts and substantially more economic than the commercial media. However, the method has so far only been applied for ^{15}N labeling (*17*) and its applicability for advanced NMR experiments and the analysis of challenging proteins has not been demonstrated.

This thesis has been aimed at the development of an economic approach to uniform isotope labeling in higher eukaryotes based on the supplementation of isotope-labeled cellular extracts as the source of amino acids in amino acid-depleted dropout media. In particular, the research work should

(i) establish a method for robust expression and high levels of isotope incorporation in insect cells
(ii) establish a step-by-step protocol for implementation in a standard laboratory environment to facilitate widespread use
(iii) provide a proof of successful application to advanced NMR experiments
(iv) establish a method for efficient deuteration in insect cells.

Organization of the following sections

The next section of this introduction (Chapter 1.2) gives an overview of isotope labeling approaches in insect cells, their development and applications. This is followed (Chapter 1.3) by a detailed description of the insect cell system, its use for recombinant protein production, and the baculovirus expression vector system. The concluding section (Chapter 1.4) discusses important aspects of the acid metabolism in insect cells and their implications for isotope labeling.

1.2 Isotope labeling in insect cells

Amino acid type selective labeling in insect cells

Amino acid type selective (AATS) labeling is well-established in insect cells and often the method of choice for NMR studies due to moderate costs and a minimum effort in preparation of the labeled growth medium (*18*, *19*). Furthermore, reduced metabolic scrambling in ICL is considered advantageous for AATS labeling (*20*), as dilution of the isotope label is significantly lowered compared to *E. coli*. So far, two approaches to AATS labeling have been established based on supplementation of the labeled amino acid to a dropout medium deficient in this amino acid or addition of the labeled amino in large excess to the full medium (excess labeling).

Excess labeling has been reported to yield incorporation of up to 84 % when ^{15}N-leucine was directly added to an otherwise unlabeled commercial medium (*19*). However, incorporation levels strongly depend on the amino acid content of the commercial medium, which is not disclosed and hence requires a laborious optimization of the supplementation scheme. A major drawback of excess labeling is the high amount of labeled amino acid, which is required for efficient isotope incorporation. Based on the desired labeling strategy, the costs of the used amino acids may become prohibitive. Therefore, excess labeling is often not feasible and replacement of individual unlabeled amino acids by labeled ones using a dropout medium is advantageous.

Initially, dropout media for AATS labeling have been prepared from several components according to established formulations (*21*). Today, commercial suppliers offer customized media at affordable prices, which remove the need for a time-consuming preparation step. Replacement of a single or several amino acids by the desired isotope-labeled amino acid(s) has been shown to yield very high incorporation rates (up to 94 %) (*18*, *19*). Typically, essential amino acids (F, I, K, L, M, T, V, and W) are preferred for this approach due to the absence of metabolic scrambling. Moreover, culture media contain considerable sources of unlabeled amino acids in the form of yeast extract and fetal bovine serum. Therefore, AATS labeling is commonly used in insect cells adapted to growth in serum-free media containing the minimal amount of yeast extract required to promote proliferation and expression.

One of the first proteins to be selectively isotope labeled in the BEVS for NMR studies was the eponymous bovine rhodopsin of the rhodopsin-like G protein coupled receptor family A. Great effort was made to prepare a selectively labeled rhodopsin sample using [α,ε-$^{15}N_2$]-lysine (*22*). This approach was successfully applied to investigate the Schiff-base linkage between retinal and the lysine residue 296 by solid state NMR. Shortly after, the first comprehensive study of AATS labeling in insect cells established incorporation of seven ^{15}N-labeled amino acids into glutathione-S-transferase (GST) (*23*). Additional insights into the amino acid metabolism in ICL were gained and used to derive a preliminary metabolic network describing scrambling of ^{15}N. NMR analysis of selectively ^{15}N-labeled GST revealed significant scrambling of amino acids related to the TCA cycle such as glutamine, glutamic acid, aspartic acid and asparagine as well as alanine. Expression media were supplemented with unlabeled yeast extracts, which are a major source of unlabeled amino acids but otherwise were required to obtain high protein yield. Consequently, additional quantities of ^{15}N-labeled amino acids were used to limit label dilution. However, a thorough analysis of the isotope incorporation was not reported.

Almost simultaneously, Strauss and coworkers reported AATS labeling of the human Abelson tyrosine kinase domain (cAbl) by ^{15}N-labeled glycine, tyrosine, phenylalanine and valine (Figure 1.1) (*18*). Remarkably, very high levels of ^{15}N incorporation (≥90 %) were achieved for the three essential amino acids, whereas significant dilution of the label was observed for glycine. High expression levels in the presence of the kinase-specific inhibitor imatinib (*24*) and the available assignment of the protein backbone (*25*) render cAbl an ideal model for optimization of labeling protocols in insect cells. Indeed, AATS labeling in the BEVS was significantly improved by modification of the previous labeling procedures (*18*, *23*) and applied to cAbl (*19*). Key to this improvement was supplementation of defined amounts of yeast extracts.

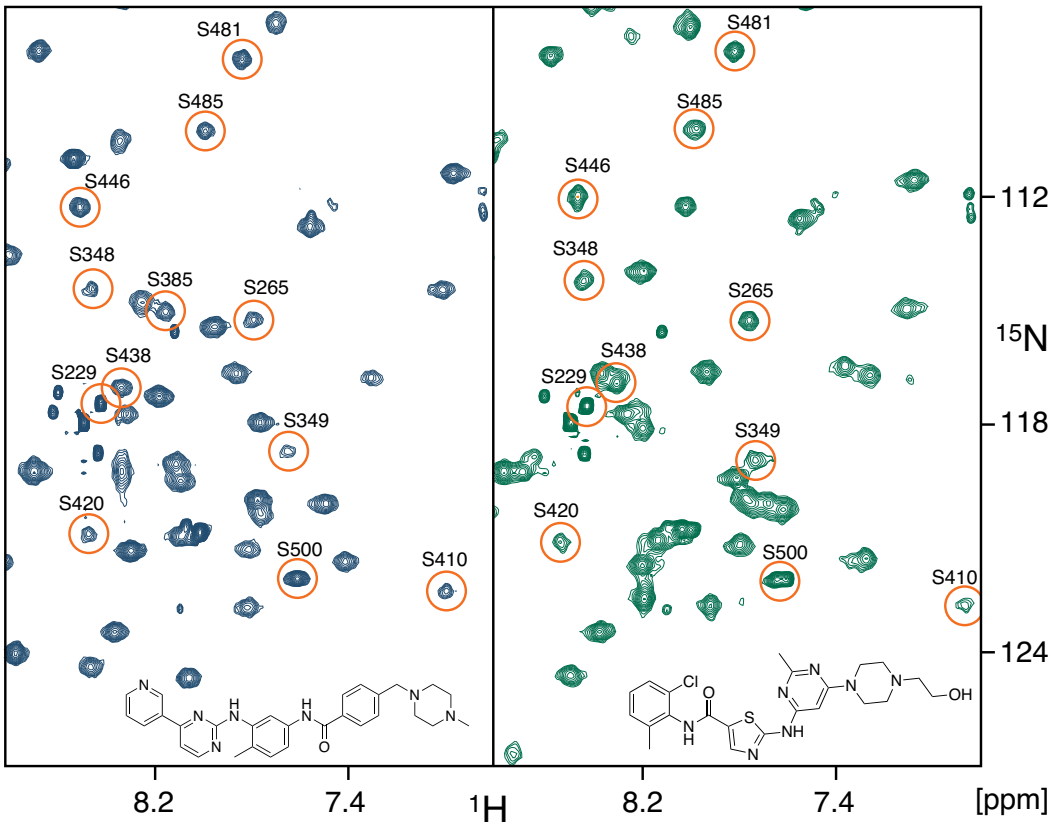


Figure 1.1 Amino acid type selective labeling of the Abelson kinase domain (cAbl) in insect cells. Supplementation of ^{15}N-labeled FGMY was according to the protocol established by Strauss and coworkers (***18***). ^{1}H,^{15}N-HSQC spectra were recorded on selectively-labeled cAbl in complex with imatinib (blue) and dasatinib (green). Interestingly, ^{15}N-serine resonances were observed in these spectra (orange circles) due to significant metabolic scrambling of glycine. The figure is adapted from (*26*).

Side chain labeling of methyl groups in methionine residues has been established as a valuable alternative to selective labeling of the protein backbone (*27*). Methyl groups exhibit a C3 rotational symmetry and display very rapid rotational motion, which results in beneficial relaxation properties for NMR studies of large proteins (*1*). Due to their conformational freedom, methionine side chains are often part of ligand binding surfaces (*28*). Selective labeling by [*methyl*-^{13}C]-methionine was first reported for the β_2-adrenergic receptor to investigate conformational changes upon agonist binding (*29*). NMR revealed changes in the

equilibrium populations for the methionine residue 82 located in the transmembrane helix 5, which may correspond to modulations of the signal transduction level in different ligand-receptor complexes.

In general, AATS labeling of essential amino acids in ICL results in high incorporation and negligible isotope scrambling due to significantly constrained amino acid metabolism (*20*, *23*). Conversely, non-essential amino acids such as glutamate and alanine are important for the cellular nitrogen metabolism (*30*) and display substantial scrambling. Furthermore, selective labeling only provides limited spectral information and often requires laborious site-directed mutagenesis to enable NMR resonance assignment (*29*). Combinatorial labeling by ^{13}C and ^{15}N has enabled backbone assignment (*27*) in selectively labeled samples and provides an alternative to the laborious mutagenesis approach.

Uniform isotope labeling in insect cells

The BEVS is not amenable to the use of inexpensive minimal media, which have been applied very successfully to uniform isotope labeling in *E. coli*. Due to the limited amino acid metabolism in ICL, growth is not possible on a single carbon source and ammonium alone. Therefore, supplementation of all canonical amino acids is required and is a key challenge to establish uniform labeling in insect cells and other higher eukaryotic expression systems such as avian and mammalian cells. Moreover, growth in media containing high fractions of heavy water (D_2O) is essential to produce uniformly deuterated proteins in *E. coli* but has not yet been established in ICL. In principle, uniform labeling can be achieved by preparation of customized media comprising isotope-labeled amino acids and their precursors, namely carbohydrates. However, as growth media for insect cells contain gram quantities of amino acids (*21*), this approach is very expensive .

An important advance for application of the BEVS in NMR has been the development of commercial media formulations providing uniform isotope labeling (BioExpress-2000, Cambridge Isotope Laboratories). Still, the use of these media is very costly, e.g. 15000 EUR per liter of ^{13}C, ^{15}N-labeled culture. Application of BioExpress-2000 was pioneered for cAbl expressed in Sf9 cells (*13*) and successfully used to assign the protein backbone to near-completeness (96 %) (*25*). This was followed by a detailed study based on residual dipolar couplings of the solution conformations of the kinase domain in complex with several inhibitors (*26*). However, more elaborate triple-resonance backbone correlation experiments on cAbl were not possible due to the large size of the protein (34 kDa) and the lack of deuteration.

Additional complications for efficient isotope labeling in the BEVS arise from the associated handling procedures. First, recombinant baculovirus stocks are commonly produced in full growth medium, which present a source of unlabeled amino acids. High-titer virus stocks are therefore essential to decrease the required volume of the viral inoculum, thereby reducing the carry-over of this unlabeled medium (*19*). Moreover, insect cells need to be adapted to growth in serum-free medium as fetal bovine sera are a source of free amino acids. High incorporation of isotopes necessitates complete removal of the unlabeled growth medium prior to expression of the recombinant protein (*18*). Medium exchange is typically achieved by centrifugation and subsequent resuspension of the cells in freshly prepared labeling medium. This is a crucial step and great care has to be taken to minimize cell damage, which will eventually lower the expression yield.

In-cell NMR studies on baculovirus-infected insect cells

In-cell NMR has gained increased momentum over recent years. Studies in living cells can provide new insights into biological function and the role of the cellular environment. *In situ* NMR studies in baculovirus-infected Sf9 cells were pioneered by Hamatsu and coworkers (*31*). Uniform isotope labeling by ^{15}N and ^{13}C was based on the commercially available expression medium BioExpress-2000 (Cambridge Isotope Laboratories). The unique property of the BEVS to suppress host cell gene synthesis and simultaneously produce high amounts of the recombinant protein under control of strong promoters in the very late phase of infection was highly advantageous. Efficient isotope incorporation (~98 %) was possible by transfer of infected Sf9 cells into the labeling medium 24 hours post infection and continued expression for another 24 hours. The utility of this approach was demonstrated for different model proteins of variable size comprising up to 148 residues. Triple-resonance backbone correlation (HNCA, HN(CO)CA, and HNCO) and ^{15}N-edited NOESY experiments were recorded on the 57 amino acid *Streptococcus* protein G B1 domain (GB1). However, non-uniform sampling and maximum entropy reconstruction of the obtained spectra were required as stability of the recombinant proteins was limiting the effective measurement time due to a decrease in the viability of the cells in the NMR tube.

Economic approaches to uniform isotope labeling in insect cells

In recent years, the development of alternative labeling methods based on cellular extracts of yeast has gained increased momentum (*16*, *17*). Commonly, these extracts are supplemented to a commercial dropout medium devoid of all amino acid sources. Remarkably, the underlying concept was reported for mammalian cell lines about two decades ago (*32*), but has not been followed up until recently. This approach is significantly more economic as compared to the commercial labeling media and enables uniform labeling by ^{15}N and ^{13}C, albeit not demonstrated in the original protocol (*16*). Moreover, supplementation of yeast extracts is well-known to promote growth and expression in insect cells (*33*).

Growth of yeast is well established in minimal media consisting of the simplest precursors for biosynthesis of all amino acids such as glycerol, methanol or glucose and ammonium salts. Preparation of yeast cell extracts requires degradation of the initial cellular biomass or the isolated protein fraction of these cells to obtain free amino acids. Two main procedures to achieve such degradation have been used to different extent, namely acidic hydrolysis at elevated temperatures (*34*) and autolytic cell digestion (autolysis) by endogenous enzymes (*16*).

The highest amount of free amino acids is generally obtained based on hydrolysis in methanesulfonic acid at 115 °C for several hours, which is used to degrade the purified protein fraction (*34*). However, a major disadvantage of this approach is the loss of amino acids by chemical degradation caused by these harsh conditions (*32*, *34*). In particular, oxidation of tryptophan and cysteine occurs, while at the same time glutamine and asparagine are converted to glutamic acid and aspartic acid, respectively. Deficiency in the non-essential amino acids cysteine, glutamine and asparagine is largely compensated for by the metabolic capabilities of the insect cell (*35-37*). Conversely, tryptophan cannot be synthesized and represents an essential amino acid (*38*).

Autolysates of yeast cells can be prepared in a laboratory environment using a bioreactor setup (*16*). Commonly, autolysis is performed under slightly acidic conditions (pH 6) and temperatures of 50 °C for several hours up to days, which enables recovery of practically all amino acids required for growth of insect cells. Consequently, autolysis is preferred over acid hydrolysis for the preparation of yeast extracts. However, for some amino acids (*20*) concentrations are very low and may become a limiting factor during expression. Preparation of homemade yeast extracts and their use for uniform isotope labeling was demonstrated earlier by expression of the human histamine receptor H1R in insect cells (*16*, *17*). However, efficient incorporation of ^{15}N (~80 %) required repeated passaging in the labeled growth medium. Furthermore, additional amino acids had to be supplemented to the medium at concentrations of up to 20 % typically used in the IPL-41 formulation (*21*) to obtain high-level expression. Most importantly, a proof of applicability has not been provided for canonical NMR experiments.

Current developments in economic isotope labeling protocols

While the work on this thesis was ongoing, further approaches to uniform labeling in ICL have been published based on cellular extracts of algae and yeast (*39*, *40*). These methods have been addressed to some extent the shortcomings of the originally reported protocol (*16*) and progress has been made towards more robust protocols for uniform labeling.

Commercial extracts of isotope-labeled yeast (CortecNet) were used for production of ^{15}N-labeled proteins in Sf9 and S2 insect cell lines (*40*). Interestingly, the protocol employs an ammonium salt to replace glutamine in the culture medium, thereby further reducing the costs. So far, this approach resulted in ^{15}N incorporation levels of up to 66 %. However, the authors suggested that starvation further increases incorporation to 80 % based on preliminary trials. Whereas, uniform labeling by ^{13}C has been proposed on the basis of this protocol, the use of deuterated yeast extracts has not been explored.

Algal extracts as a source of labeled amino acids have been pioneered in higher eukaryotic expression systems over two decades ago. Hansen and coworkers used cellular extracts of algae obtained by acidic hydrolysis to establish uniform isotope labeling in mammalian cell lines (*32*). Nowadays, amino acid mixtures prepared from algal extracts are available from different commercial sources (Cambridge Isotope Laboratories and Sigma-Aldrich) providing all commonly applied labeling schemes at minimal costs, which may enable widespread application of these products. Whereas yeast cells still require organic carbon sources such as glycerol, methanol or glucose, algal cells assimilate CO_2 to form complex carbohydrates. Therefore, the use of algal extracts is particularly cost-effective for uniform ^{13}C labeling.

The first protocol for uniform labeling by ^{15}N, ^{13}C and ^{2}H based on commercial algal extracts (*39*) has been established shortly after the initial application of commercial yeast extracts (*40*). This approach involves supplementation of 5 mM ammonium salt to compensate for deficiencies of glutamine and asparagine in the algal extract similar to the labeling protocol based on commercial yeast extracts (*40*). Supplementation of the essential amino acid ^{15}N-labeled tryptophan has been required but at significantly reduced concentrations as compared to commercial media formulations. In general, incorporation levels of 80 % and 75 % were achieved for uniform ^{15}N and ^{13}C labeling, respectively. Moreover, partial deuteration with overall incorporation levels of up to 76 % was reported for labeling by ^{2}H,^{15}N and ^{2}H,^{13}C,^{15}N. However, incorporation of deuterium was only estimated

from the total isotope shift in combination with ^{15}N or ^{13}C, ^{15}N labeling by mass spectrometric analysis, but a direct analysis of the deuteration level was not carried out.

Earlier, the use of commercial algal extracts in ICL has been reported in combination with AATS labeling of the β_2-adrenergic receptor (β_2AR) by [$\alpha\beta\gamma$-^{2}H, *methyl*-^{13}C]-methionine (*41*). This tailor-made labeling scheme has been utilized to provide fractional background deuteration based on highly deuterated algal amino acid mixtures, thereby increasing the sensitivity of the investigated methionine side chain resonances. Incorporation of ^{2}H into non-exchangeable backbone and side chain protons as high as 90 % has been obtained for fourteen labeled amino acids using a meticulous protocol for addition of individual amino acids. Remarkably, a lowered overall deuteration has been found for non-essential amino acids, which is caused by metabolic scrambling in insect cells. The protocol enabled to study dynamics of β_2AR in lipid bilayers, which displayed significant differences to the detergent-solubilized receptor (*29*). However, broader application of this protocol is hindered by the laborious labeling procedure, which requires the costly supplementation of several deuterated amino acids at defined time points during expression.

Considerations of the medium osmolarity

A common limiting factor of the labeling approaches is the maximum osmolarity tolerated by insect cells. Commercial media are typically adjusted to an osmolarity range of 320 to 380 mOsm/kg resembling the physiological conditions in the insect hemolymph, which are optimal for growth and protein expression (*42*). The commercial media typically contain about 4 g/L yeast extract (*43*) corresponding to 50-70 mOsm/kg. For isotope labeling dropout media depleted in all amino acid sources are used with an osmolarity of about 220 mOsm/kg. Supplementation of such a dropout medium with 4 g/L yeast extract increases the osmolarity to 270-290 mOsm/kg and provides amino acids at an overall concentration of about 1.5-2 g/L (*20*). This amino acid concentration is however not sufficient to promote growth and expression in insect cells. As the optimal osmolarity is not yet reached, additional amounts of labeled extracts may be supplemented to increase the amino acid content. Indeed, high expression levels have been obtained by supplementation of yeast and algal extracts at concentrations of 10 g/L (*39*, *40*).

1.3 Recombinant expression in insect cells

Stable transfection of insect cells

Several commercial systems are available for stable transfection, which enables long-term integration of heterologous genes into the ICL and constitutive expression of the recombinant protein. Stably transfected cells will efficiently propagate these genes to their progeny.

Stable transfection is most widely used in the lepidopteran cell line Sf9 (*44*) and Schneider 2 (S2) cells derived from late embryonic stages of *D. melanogaster* (*45*). Recombinant expression in stably-transfected S2 cells is commonly based on the actin and the metallothionein promoters of *D. melanogaster* (*46*, *47*). A commercial *Drosophila* Expression System (DES) was established based on the metallothionein promoter, which offers tight gene regulation and efficient selection by an accessory vector mediating antibiotic resistance (*48*). Several proteins have been expressed in milligram quantitates using the commercial DES (*49*).

In lepidopteran cell lines, such as Sf9 cells, stable transfection is based on the use of immediate early promoters found in different baculoviruses (*50*, *51*). This approach was commercialized in the InsectSelect™ and InsectDirect® system. Selection of a suitable clone for constitutive expression is similar to the DES, but may further be accomplished by incorporation of the resistance gene cassette into a single expression vector. Moreover, milligram amounts of secreted and functional active protein were obtained per liter of stably transfected cell culture (*52*).

Unfortunately, stably transfected insect cell lines often display significant differences in the expression level related to the copy number of the gene and unspecific integration. Therefore, clonal selection of single cells is used to determine the optimal clone for protein production. However, this step is laborious due to the involved screening effort and the lengthy scale-up process starting from a single cell. Site-specific integration of heterologous genes was established to facilitate optimization of recombinant expression following stable transfection. Based on random integration of gene cassettes comprising a reporter and resistance gene, the screening process is greatly accelerated. These gene cassettes are flanked by recombinase recognition sites, which enable integration of the heterologous gene by recombinase-mediated cassette exchange following co-transfection with plasmids containing the recombinase gene and the gene of interest (*53*). This approach has been pioneered in Chinese Hamster Ovarian (CHO) cells (*54*) and was recently expanded to Sf9 cells (*55*).

A key limitation of stably transfected ICL is the requirement for an extensive selection process, which also involves several weeks of scale-up procedure. Furthermore, the overall yield of recombinant protein is typically lowered compared to the baculovirus expression vector system.

The baculovirus-infected insect cell system

Pioneering work that laid the foundation of the baculovirus expression vector system (BEVS) reaches back to the 1980s (*56*), when the viral polyhedrin protein was shown to be non-essential for propagation of the baculovirus in cell cultures. The potential for recombinant expression of heterologous proteins was immediately realized and used to obtain high amounts of the human interferon beta (*57*). Since then the baculovirus/insect cell system has seen major improvements including elaborated tools for genetic manipulation of the baculovirus as well as the ease of culturing (*58*) and eventually emerged as the predominantly

used eukaryotic expression system (*3*). Very strong viral promoters in the BEVS enable high yields of recombinant proteins, which often surpass the productivity obtained by stably transfected insect cells. Consequently, due to increased expression levels and the reduced timespan of the process the use of the BEVS is often favored over stable transfection of ICL.

Recombinant expression based on the BEVS is well-established in the lepidopteran cell lines Sf9 and Sf21 originating from the IPLB-SF-21 cell line, which was isolated from ovarian tissue of *Spodoptera frugiperda* pupae (*44*). Sf9 cells were established as a clonal isolate of the Sf21 cell line. Moreover, the use of the BEVS in High-Five™ cells originally derived from BTI-TN-5B1-4 ovarian cells of the cabbage looper *Trichoplusia ni* (*59*) has seen wide application, especially for highly efficient secretion of glycoproteins. All of these cell lines support growth in serum-free medium as suspension or adherent cultures and propagation of the recombinant baculovirus. The highest productivity of the BEVS has been obtained in High-Five™ cells (*60*, *61*), but a significant decrease in expression levels was reported at higher cell densities (*62*). In addition, maintenance of this cell line in suspension is challenging, which is related to an increased propensity to form cell aggregates. Therefore, efficient generation of recombinant baculovirus is preferentially performed in Sf9 and Sf21 cell lines (*7*, *20*).

Recent approaches have established intact silkworm pupae and larvae as expression hosts to obtain milligram quantities of purified proteins using the BEVS (*63*). Furthermore, expression of uniformly ^{15}N-labeled immunoglobulin G for NMR studies was reported for larvae reared on a diet containing isotope-labeled yeast extracts (*64*).

The baculovirus life cycle and heterologous gene expression

The baculovirus *Autographa californica* multiple nucleopolyhedrovirus (*Ac*MNPV) was originally isolated from the alfalfa looper *Autographa californica* (*65*). Two structurally distinct states of *Ac*MNPV were described, namely the highly infectious budded virus (BV) and the polyhedral-shaped occlusion-derived virus (OV) (*66*, *67*). Infection of insect cells by parental baculovirus (Figure 1.2) involves a regulatory cascade consisting of four gene classes related to expression of viral proteins: early, delayed early, late and very late genes. The class of very late genes is characteristic to baculoviruses and of great importance for recombinant expression in insect cells based on the BEVS. Initial attachment to the host cell occurs during the early phase of the infection cycle and is followed by the release of the viral DNA into the host cell, effectively suppressing cellular gene expression (*68*). Viral entry is accomplished by clathrin-mediated endocytosis in less than an hour. Within the first 10 to 20 hours post infection (hpi) viral DNA is replicated in the nucleus of the infected host cell and further enveloped by the cellular plasma membrane to form new BV. This is referred to as the late phase of infection and involves exponential generation of these virions, which are consequently released to propagate baculovirus infection throughout the host known as a secondary wave of infection (*69*). A characteristic decrease in the protein synthesis of the host cell is observed about 18 hpi and associated with very high viral titers (*70*). During the very late phase (24-48 hpi) of the infection cycle occlusion bodies are formed containing large numbers of enveloped virions, which are further embedded in a protein matrix comprised of polyhedrin. This protein is expressed under the control of the very strong *polh* promoter and was shown to require the host factor polyhedrin promoter binding protein acting in *trans* (*71*). Moreover, fibrillary structures are observed in the cytoplasm and the nucleus of the host cells formed by the p10 protein. Expression of the *p10* gene is driven by

the second very late promotor *p10* (*72*, *73*) and is crucial for efficient release of occlusion bodies (*74*). Concurrent production of both proteins in very high amounts is specific to the very late phase. However, propagation of the baculovirus coupled to the generation of progeny virions is limited to the late phase of the infection. Therefore, both viral genes are amenable to substitution by heterologous genes to provide high-level expression of recombinant proteins in the very late phase. Remarkably, *polh* and *p10* promoters can drive heterologous gene expression at different loci in the viral genome and allow for multiple copies to increase the yield of recombinant protein.

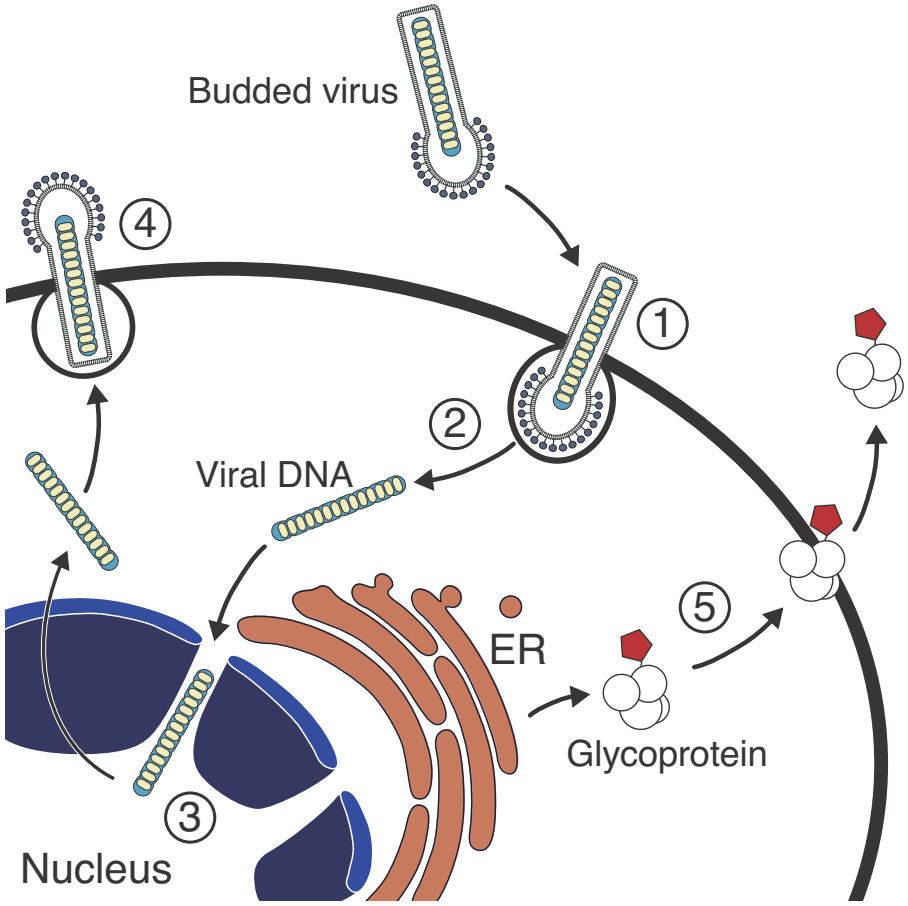


Figure 1.2 Principle of the baculovirus expression vector system (BEVS) in insect cells. Entry of the budded virus into the host cell is accomplished via clathrin-mediated endocytosis (1) and immediately followed by the release of the virion nucleocapsid (2). Uncoating of the viral DNA occurs in the nucleus of the cell to initiate replication in the late phase of the infection (3). Subsequently, release of newly formed budded virus is triggered (4), which results in a secondary infection wave infecting adjacent insect cells. In the very-late phase of infection host cell gene synthesis is shut down. Simultaneously, very-late and strong viral promoters drive expression of the recombinant protein. Glycoproteins are routed via the secretion pathway of the host cell and efficiently secreted to the medium (5).

An important step towards the development of the BEVS was the observation that transfection with viral DNA alone could cause baculovirus infection (*75*). Today, recombinant baculovirus DNA (bacmid) is commonly generated by *in vivo* transposition in *E. coli* (*76*) or recombination in insect cells (*77*). This bacmid is used for transient

(co-)transfection of insect cell lines to obtain recombinant baculovirus in adherent and suspension cultures, followed by up to three amplification steps to yield high titer virus stocks.

Limitations of the baculovirus expression vector system

The metabolic capabilities of insect cell cultures are significantly reduced as compared to *E. coli* or lower eukaryotes such as yeast and algae. Consequently, complex media providing several essential components are required to promote growth. However, a wide array of commercial medium formulations including serum-free and protein-free media is available at minimal cost. ICL are commonly maintained as permanent cultures for several passages, which requires sub-culturing routines in the order of days. Recombinant expression in insect cells usually varies from two to four days and thus compares less favorably with the faster bacterial or lower eukaryotic expression systems.

A crucial step in the overall process of the BEVS and unique to baculovirus-infected insect cells is the requirement to generate and amplify recombinant baculovirus. Major progress was made to simplify the construction of recombinant viral DNA. Artificial baculovirus genomes (bacmid) were established, which are constitutively present in *E. coli* (*76*). *In vivo* transposition based on Tn7-recombinase in the engineered *E. coli* DH10Bac™ strain is used to introduce the recombinant gene into a parental bacmid. Following clonal selection, the recombinant bacmid can be isolated and enables transient transfection of ICL. To date, several commercial solutions are available, which combine the original bacmid technology with novel cloning approaches such as the GATEWAY system (Life Technologies). Recently, expression of protein complexes comprising multiple subunits was established in the BEVS by the MultiBac system (*78*). Homologous recombination is used to generate a single baculovirus expression vector, which may provide the correct stoichiometry of the subunits for these protein complexes. Moreover, multilocus vectors are available for expression of several large proteins by a single recombinant virus (*79*). Irrespective of this technological development, generation of recombinant baculovirus remains a highly time-consuming step in the BEVS.

Following successful transfection of insect cells with recombinant bacmid DNA, high amounts of budded baculovirus are released to the culture supernatant in the late phase of infection. High titer stocks for expression cultures are obtained in consecutive amplification rounds by infection of insect cells with this supernatant. These stocks are typically stable for several months and enable repetitive protein production. The additional amplification of the recombinant baculovirus stock can effectively be avoided by the use of baculovirus-infected insect cells (BIICs), which allow for titer-independent and reproducible infection at large-scale (*80*).

The BEVS utilizes strong viral promoters to obtain remarkably high expression levels for cytosolic proteins that often exceed the capacity of mammalian cell lines such as HEK293 or CHO cells. However, production of secreted glycoproteins and transmembrane proteins is often significantly lower compared to cytosolic proteins (*81*). This is not only related to routing or post-translational processing of these proteins, but further due to the progressing infection of the recombinant baculovirus (*81*). At the very late phase of the infection cycle, the viral chitinase (*chiA*) is expressed in high amounts, which is required to degrade the chitin skeleton for the release of occlusion bodies (*82*). Accumulation of this enzyme in the endoplasmic reticulum (ER) (*83*) may interfere with the secretion of the recombinant

glycoprotein (*84*). Moreover, high levels of the viral cysteine protease cathepsin are observed towards the end of the infection, which also predominantly localizes in the ER (*85*). Cathepsin accumulation is likely to impede secretion and may further result in degradation of the glycoprotein. Baculovirus expression vectors deficient in the genes encoding for the viral chitinase and cathepsin have been described to increase the yield of secreted proteins (*86*). Interestingly, expression of *chiA* is essential to obtain functional cathepsin (*87*), hence deletion of this gene is sufficient to suppress viral proteolytic activity. As a consequence, commercial BEVS have been established comprising a deletion in the *chiA* gene to enhance recombinant expression of glycoproteins.

Whereas glycosylation is present in ICL, it is limited to simple oligo-mannose sugar chains. Mammalian cell lines exhibit complex glycosylation patterns, which involve terminal sialylation (*53*, *88*). Simple glycosylation patterns are well suited to provide homogenous samples of diffracting quality for X-ray crystallography. Conversely, alterations in the glycosylation pattern may impede folding and compromise biological activity of mammalian proteins (*89*, *90*). Several stably transfected ICL have been established to overcome this limitation. In particular, stable transfection has been used to include sialylation and galactosylation by heterologous genes encoding a specific sialyltransferase and the β-1,4 galactosyl transferase of mammalian origin (*91*, *92*). Although these transformed cell lines have been shown to produce sialylated glycosylation patterns, the actual benefit for the BEVS is limited, as the use of immediate early viral promoters eventually shuts down the expression of these transferases in the very late phase of the progressing infection. Moreover, protein inhomogeneity may increase as described for recombinant expression in mammalian cells, which often hampers the crystallization process of these proteins.

In contrast to other widely used expression systems including mammalian cell lines, screening of constructs prior to large-scale expression is laborious in the BEVS. One way around this limitation is the transient transfection of insect cells. However, the expression levels obtained by transiently transfected cells may not necessarily correlate with the productivity of infected cells based on the BEVS. Another approach to enable high-throughput expression using the BEVS relies on the conventional bacmid generation and a further *in vivo* recombination step. Successful recombination rescues a non-viable gene deletion, thereby selecting for the recombinant baculovirus (*93*). The method allows for integration into a robotic setup and rapid isolation of parental baculovirus. Commonly used commercial systems based on this technology are referred to as BacMagic (Merck) and flashBAC (Oxford Expression Technologies). The latter system is often utilized to generate recombinant baculovirus as an alternative to *in vivo* transposition in *E. coli*.

1.4 Aspects of the amino acid metabolism in insect cells pertinent to isotope labeling

Essential amino acids and the nitrogen metabolism

Insect cells lack the ability to synthesize isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan and valine (*94*). Hence, these amino acids are not subjected to metabolic scrambling. This is highly advantageous for selective labeling approaches, which are less affected by label dilution as compared to *E. coli*. Conversely, insect cells can synthesize all remaining canonical amino acids mainly through activity of different amino transferases and utilization of the carbon scaffold of glucose, which is metabolized via glycolysis and the TCA cycle. Moreover, important methyl transfer reactions are involved in metabolic scrambling of glycine and biosynthesis of cysteine. These metabolic pathways are highly relevant to isotope labeling (Figure 1.3).

The nitrogen metabolism in insect cell lines (ICL) has been extensively studied revealing a strong interrelation between the carbon and the nitrogen metabolism (*35*, *37*, *95*). In general, several metabolic pathways are identical to the ones described for mammalian cell cultures. However, insect cells display a remarkable difference in the assimilation of nitrogen depending on the available nutrients.

The central role of glutamate, aspartate and alanine

Glutamate is commonly formed from α-ketoglutarate in a reversible reaction catalyzed by glutamate dehydrogenase. However, a second metabolic pathway for biosynthesis of glutamate in insect cells has been reported and involves reductive amination of α-ketoglutarate by action of the NADH-dependent glutamine synthetase/glutamate synthase (GOGAT) (*35*, *95*). Apparently, this is the preferred pathway for nitrogen assimilation in ICL. Regulation of these pathways is dependent on the extracellular concentrations of glutamine and glucose (*35*). Most notably, glutamate plays a key role in several transamination reactions, namely biosynthesis of alanine, aspartate and serine and is the direct precursor for the synthesis of glutamine. Formation of alanine and aspartate is achieved via a reversible transamination reaction step from the α-keto acids pyruvate and oxaloacetate, respectively (*30*). Alanine is a generic metabolic product in insect cells and predominantly formed under glucose excess. Conversely, limitation of glucose results in secretion of ammonium into medium, which hints on increased gluconeogenesis and anaplerotic flux of glutamine into the TCA via the α-ketoglutarate node by consecutive activity of glutaminase and glutamate dehydrogenase (*35*). Hence, a decrease is observed in the transamination rate of glutamate to synthesize alanine from pyruvate.

Biosynthesis of asparagine, glutamine and proline

Remarkably, synthesis of asparagine and glutamine from their respective precursor aspartate and glutamate involves two distinct pathways. Glutamine is formed by direct condensation of ammonium with glutamate by the activity of glutamate-ammonia ligase (GAL), whereas biosynthesis of asparagine by aspartate-ammonia ligase synthetase (AAL) requires the transfer of the amide group of glutamine to aspartate (*32*).

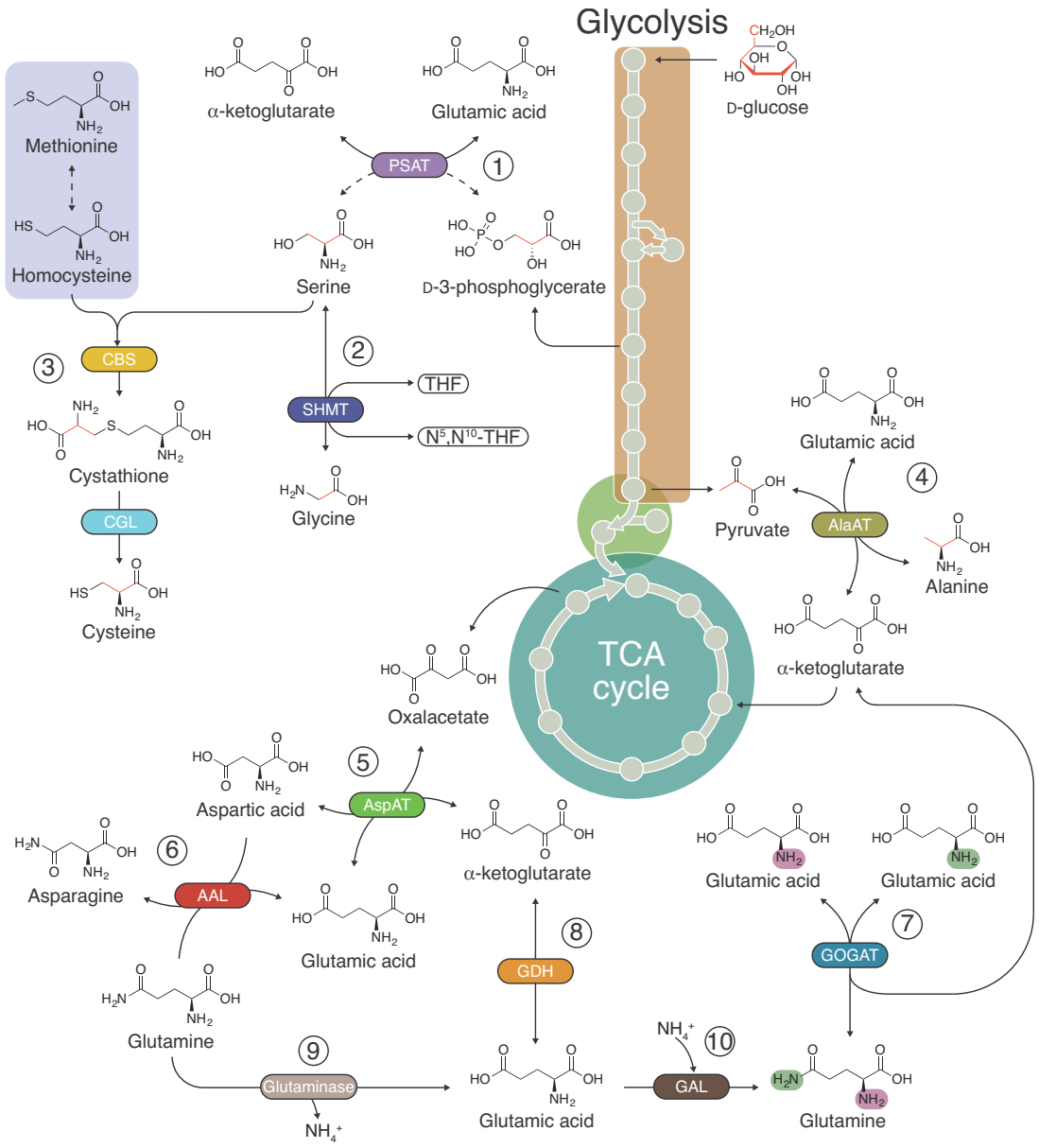


Figure 1.3 Metabolic scrambling of amino acids in insect cell lines. Pathways involving more than one reaction step are simplified (dashed arrows). Enzymatic reactions are as follows: 1, transamination of D-3-phosphoglycerate by phosphoserine aminotransferase (PSAT); 2, conversion of serine to glycine by activity of serine hydroxymethyltransferase (SHMT); 3, reverse transsulfuration pathway to form cysteine by consecutive activity of cystathione beta synthase (CBS) and cystathione gamme-lyase (CGL); 4, transamination of pyruvate by alanine amino transferase (AlaAT); 5, transamination of oxaloacetate by aspartate amino transferase (AspAT); 6, formation of asparagine by aspartate-ammonia ligase (AAL); 7, incorporation of glutamine amide and amino nitrogen into glutamate by activity of NADH-dependent glutamine synthetase/glutamate synthase (GOGAT); 8, reversible deamination of glutamate by glutamate dehydrogenase (GDH): 9, conversion of glutamine to glutamate by glutaminase and 10, condensation of ammonium with glutamate by glutamate-ammonia ligase (GAL). Transamination (1, 2, 4, and 5) and reactions of GDH and GOGAT are generally reversible (double arrows). Glycolytic scrambling of the glucose carbon scaffold is indicated in red. The unique transfer of amino and amide nitrogen catalyzed by GOGAT to form two molecules of glutamate is color-coded in purple and green, respectively.

Further, biosynthesis of proline is based on spontaneous cyclization of glutamate-5-semialdehyde, which is formed from glutamate by two consecutive reactions via activity of glutamate-5 kinase and glutamate-5-semialdehyde dehydrogenase (*38*). This cyclization will result in exchange of the H^{δ} positions and may decrease incorporation of deuterium. Moreover, unlabeled glucose is scrambled into synthesized proline in form of the carbon backbone of glutamate, which exits the TCA at the α-ketoglutarate node.

Metabolism of serine, glycine and cysteine

Biosynthesis of serine, cysteine and glycine is strongly interrelated in all animal cell lines (*38*) and is responsible for substantial metabolic scrambling of these amino acids. Serine is derived from 3-phosphoglycerate, an intermediate of the glycolytic degradation of glucose. Hence, the carbon scaffold of glucose is transferred to serine. This pathway involves a notable transamination step catalyzed by phosphoserine transaminase, which again utilizes glutamate as the amino group donor (*38*).

Serine serves as a methyl donor for covalent methylation of tetrahydrofolate to generate N^5, N^{10}-methylen tetrahydrofolate (N^5, N^{10}-THF) by serine hydroxymethyl transferase, which will result in the formation of glycine (*38*). N^5, N^{10}-THF is a versatile donor of C_1 carbon units in several metabolic reactions such as thymidine biosynthesis. Moreover, condensation of N^5, N^{10}-THF with CO_2 and ammonium by glycine synthase represents an alternative pathway for synthesis of glycine in animal cells (*38*). Both pathways will eventually incorporate solvent protons into H^{α} position of glycine and substantially lower the level of deuterium labeling.

Whereas methionine is considered an essential amino acid, cysteine can by synthesized *de novo* by condensation of homocysteine and serine. Homocysteine is an intermediate of the *S*-adenosylmethionine (SAM) cycle, which involves continuous demethylation and regeneration of methionine to provide an important methyl donor for various metabolic reactions such as synthesis of nucleic acids and lipids (*38*). However, homocysteine can be removed from this cyclic pathway by action of cystathionine-β-synthase to form cystathionin, followed by a final deamination and cleavage step to yield cysteine and α-ketobutyrate. Consequently, this enables biosynthesis of cysteine by transsulfuration of methionine in insect cells. As a result, cysteine comprises the carbon scaffold of the precursor serine.

Changes in the amino acid metabolism during baculovirus infection

Interestingly, baculovirus infection was shown to reduce overall influx of most of the amino acids into Sf9 cells (*96*). This decrease in amino acid incorporation may hint on a significant cellular pool of amino acids, which is utilized during growth arrest in infected cells. Consequently, dilution of the isotope label may be expected due to residual unlabeled biomass and reduced uptake of labeled amino acids from the growth medium. In contrast, production of alanine was remarkably increased during viral infection. Simultaneously, higher glucose consumption is observed and results in increased formation of pyruvate. Moreover, baculovirus-infected insect cells displayed increased activity of the GOGAT, which transfers the amide-nitrogen of glutamine to α-ketoglutarate yielding two molecules of glutamate (*95*). Glutamate is channeled into the TCA cycle via the α-ketoglutarate node, thereby liberating considerable amounts of ammonium. ICL employ substantial fractions of pyruvate to generate alanine through activity of alanine aminotransferase (AlaAT) for

ammonium detoxification. This is an important source of label dilution as pyruvate is formed by glycolytic degradation of unlabeled glucose. Indeed, significant label dilution was observed for ^{13}C-alanine in the commercial medium (*13*) and media based on algal extracts (*39*). Clearly, this metabolic pathway will further result in protonation of alanine. Hence, supplementation of ^{13}C-labeled and potentially deuterated glucose may be required to achieve high levels of isotope incorporation. Conversely, the use of L-cycloserine, as specific inhibitor of AlaAT (*97*), was sufficient to improve incorporation of labeled alanine into the expressed protein (*39*) and allowed to supply unlabeled glucose in the labeling medium.

Implications for isotope labeling

An important consequence of this strong interrelation between the carbon and the nitrogen metabolism is that high incorporation of ^{13}C further requires supplementation of ^{13}C-labeled glucose in addition to labeled amino acids. Moreover, all transamination reactions and the demethylation of serine to glycine integrate solvent protons into the H^{α} positions of the resulting amino acids, which impedes efficient incorporation of deuterium. Finally, $^{15}NH_4^+$ is incorporated into glutamine and alanine under glucose excess, whereas glucose limitation increases the metabolic flux of glutamine into the TCA cycle via the α-ketoglutarate node. All these pathways contribute significantly to scrambling of ^{15}N and involve glutamic acid as a central molecule, which simultaneously functions as an amine acceptor and a donor for transamination reactions. Therefore, supplementation of glutamine is not strictly required, when ammonium is added to the prepared labeling media (*37*) but will result in increased incorporation of glucose-derived carbon scaffolds into several amino acids via the TCA cycle.

2 Development and application of an economic approach to uniform isotope labeling in insect cells

Organization of the chapter

The first section (Chapter 2.1) is a reprint of the original publication Opitz, Isogai and Grzesiek (*98*), which describes the use of isotope-labeled yeast extract to obtain uniformly labeled proteins in baculovirus-infected insect cells. Economic implementation of the labeling procedure is presented in a step-by-step protocol. The approach is successfully applied to complex targets of human origin using advanced NMR experiments.

The following section (Chapter 2.2) reports a mass spectrometric approach to analyze the isotope incorporation in the uniformly labeled Abelson kinase (cAbl) and the green fluorescent protein (GFP) expressed in insect cells, which were grown on ^{2}H,^{15}N-labeled yeast and perdeuterated algal extracts, respectively. The method allows simultaneous analysis of different isotopes resolved for individual amino acids.

2.1 An economic approach to efficient isotope labeling in insect cells using homemade ^{15}N-, ^{13}C- and ^{2}H-labeled yeast extracts

Original publication

Christian Opitz, Shin Isogai and Stephan Grzesiek

An economic approach to efficient isotope labeling in insect cells using homemade ^{15}N-, ^{13}C- and ^{2}H-labeled yeast extracts

Journal of Biomolecular NMR, 2015 vol. 62 (3) pp. 373-285.

J Biomol NMR
DOI 10.1007/s10858-015-9954-3

ARTICLE

An economic approach to efficient isotope labeling in insect cells using homemade ^{15}N-, ^{13}C- and ^{2}H-labeled yeast extracts

Christian Opitz[1] · **Shin Isogai**[1] · **Stephan Grzesiek**[1]

Received: 28 April 2015 / Accepted: 1 June 2015

Abstract Heterologous expression of proteins in insect cells is frequently used for crystallographic structural studies due to the high yields even for challenging proteins requiring the eukaryotic protein processing capabilities of the host. However for NMR studies, the need for isotope labeling poses extreme challenges in eukaryotic hosts. Here, we describe a robust method to achieve uniform protein ^{15}N and ^{13}C labeling of up to 90 % in baculovirus-infected insect cells. The approach is based on the production of labeled yeast extract, which is subsequently supplemented to insect cell growth media. The method also allows deuteration at levels of >60 % without decrease in expression yield. The economic implementation of the labeling procedures into a standard structural biology laboratory environment is described in a step-by-step protocol. Applications are demonstrated for a variety of NMR experiments using the Abelson kinase domain, GFP, and the beta-1 adrenergic receptor as examples. Deuterated expression of the latter provides spectra of very high quality of a eukaryotic G-protein coupled receptor.

Keywords Abelson kinase · GPCR · GFP · Beta-1 adrenergic receptor · Baculovirus · Heteronuclear NMR

Introduction

Labeling by the stable isotopes ^{15}N, ^{13}C and often also ^{2}H is a prerequisite for the majority of contemporary biomolecular NMR experiments. It is well established for protein expression in the heterologous host *Escherichia coli* where minimal media provide an inexpensive, straightforward manner of isotope incorporation. However, high-quality functional expression of proteins from higher organisms such as humans often requires a higher eukaryotic host system. This seems related to the availability of a proper protein folding, post-translational modification, and membrane insertion machinery, which is less developed or absent in *E. coli*.

The baculovirus/insect cell (*Spodoptera frugiperda*) system is a well-established, highly successful eukaryotic expression system capable of many post-translational modifications, which has been used in numerous crystallographic studies of soluble and membrane proteins. However, the restricted amino acid metabolism of insect cells presents an extreme challenge for isotope labeling. Typical insect cell growth media are complex (O'Reilly et al. 1994) and require the addition of all amino acids in isotope-labeled form. Based on a commercial medium containing labeled amino acids, the uniform ^{15}N and ^{13}C labeling and high-yield expression of the human Abelson kinase domain (c-Abl) has been demonstrated in this system (Strauss et al. 2005). In combination with selective labeling of individual amino acids (Strauss et al. 2003), almost complete backbone assignments and high resolution structural information could be obtained for the 32 kDa c-Abl using this uniformly ^{15}N/^{13}C-labeled material (Vajpai et al. 2008a, b). Although this procedure of uniform ^{15}N and ^{13}C labeling is clearly feasible, it is extremely expensive. Furthermore, for the majority of larger

Electronic supplementary material The online version of this article (doi:10.1007/s10858-015-9954-3) contains supplementary material, which is available to authorized users.

✉ Stephan Grzesiek
Stephan.Grzesiek@unibas.ch

[1] Focal Area Structural Biology and Biophysics, Biozentrum, University of Basel, 4056 Basel, Switzerland

Published online: 13 June 2015

expression targets, labeling by ^{2}H would be required for optimal sensitivity and resolution. However, the costs of deuterated ^{15}N- and/or ^{13}C-labeled amino acids are prohibitive, and only addition of deuterated, but otherwise unlabeled amino acids has been reported (Kofuku et al. 2014).

As alternative to the addition of labeled amino acids to the insect cell medium the addition of labeled yeast autolysates (yeastolates) has been proposed (Egorova-Zachernyuk et al. 2009). Incorporation of ^{15}N by supplementation of such extracts as the sole amino acid source to an inexpensive dropout medium was demonstrated for the human histamine receptor H1R (Egorova-Zachernyuk et al. 2010). However, the method was not further developed for high-efficiency isotope labeling including ^{13}C and ^{2}H as well as applications to multi-dimensional NMR. A recent refinement of the yeastolate-based approach (Meola et al. 2014) achieved protein expression yields of 6–11 mg/L and ^{15}N incorporation levels between 65 and 80 % by supplementation with a labeled yeast extract from a commercial source of undisclosed composition.

Here we have developed a robust method for the production of ^{15}N- and ^{13}C-labeled yeast extract and its use as supplement in insect cell growth media for protein expression. The method achieves uniform protein ^{15}N and ^{13}C labeling of up to 90 % at typical expression yields of 40 mg/L for soluble proteins. By growing yeast in deuterated media, the procedure can also yield protein deuteration levels of >60 % without decrease in insect cell expression yield. The methods are applied to c-Abl, GFP, and the beta-1 adrenergic receptor. For the latter, ^{2}H,^{15}N labeling provides spectra of so far unprecedented quality for a eukaryotic G-protein coupled receptor. The labeling procedures are described in a step-by-step protocol for economic implementation in a standard structural biology laboratory environment.

Results and discussion

Production and characterization of optimized *Pichia pastoris* yeastolate

Detailed protocols for the yeast fermentation and the production of yeastolate are given in Table 1. In brief, *Pichia pastoris* was grown in a fermenter under fed-batch conditions with typical volumes of 0.5–5 L. Growth on a glucose- and ammonium-based yeast minimal medium (YM) yielded up to 64 g cell wet weight (CWW) of yeast biomass per liter of cell culture medium in ^{15}N- or $^{15}N/^{13}C$-labeled form. Deuteration could be achieved by adapting the yeast cells first on a 70 % D_2O-YM medium, which was followed by a switch to 100 % D_2O-YM. Under these conditions, still 48 g/L CWW were produced, corresponding to a reduction by 25 % relative to the H_2O conditions.

An optimized protocol for *Pichia pastoris* autolysis was developed, which maximizes the amino acid content in the yeast extract. Among others, the protocol contains a zymolyase incubation step to process the yeast cell wall prior to autolysis and papain treatment during the high-temperature incubation phase. The protocol yields on average 6.6 ± 0.7 g lyophilized yeastolate per liter of cell culture, containing 38 ± 4 % (w/w) of free amino acids.

The isotope incorporation levels of the free amino acids in the yeastolates were determined by LC/MS analysis (Supplementary Table S1). Using $^{15}NH_4Cl$ as the sole nitrogen source for yeast growth, an average ^{15}N incorporation of 96 % was observed. Likewise, $^{13}C_6$-glucose as the sole carbon source yielded an average ^{13}C incorporation of 97 %. Growth on 100 % D_2O-YM, but not using any other deuterated ingredients, resulted in an average ^{2}H incorporation of 75 %. This is in good agreement with previously reported deuteration levels in *E. coli* for growth on D_2O and otherwise protonated medium components (Rosen et al. 1996).

Optimization of yeast extract supplement to insect cell medium

To establish the amount of amino acids required as supplement to an amino acid-depleted, yeast extract-free insect cell medium (ΔSF4, Bioconcept), cell proliferation after baculovirus infection as well as the expression of c-Abl were assayed for Sf9 insect cells grown on this medium supplemented by varying amounts of amino acids. Starting with amino acid concentrations according to the established IPL-41 medium (Weiss et al. 1981), the supplementation was step-wise reduced. Reduction to 20 % of the IPL-41 amino acid level did not affect cell viability and protein expression, while any further reduction led to a significant decrease of both parameters.

In a next step, the content of single amino acids in a commercial (Bacto Yeast Extract, Difco) and the *Pichia* yeastolate was quantified by mass spectrometric phenylthiohydantoin amino acid analysis (PTH-AAA). Cysteine was not detectable by PTH-AAA, but an earlier study indicates a low content [<1 % (w/w)] in various yeasts (Martini et al. 1979). Figure 1 shows the determined amino acid concentrations for 8 g/L commercial and *Pichia* yeastolates in comparison to the 20 % IPL-41

Table 1 Fermentation of *Pichia pastoris* and preparation of yeastolate

Preparation of yeast medium (YM)	
Unlabeled YM	1. Prepare unlabeled minimal medium based on glucose and yeast nitrogen base (YNB) in ultrapure water (Milli-Q) as 100 mM potassium phosphate, 2.5 g/L glucose, 2 mg/L biotin, and 6.7 g/L YNB without amino acids (e.g. BD Difco)
	2. Adjust medium to pH 6
	3. Filter-sterilize and store at 4 °C
^{15}N-YM	Follow recipe for unlabeled YM, but use YNB without amino acids and without ammonium (e.g. BD Difco). Add 4.3 g/L $^{15}NH_4Cl$ as nitrogen source
^{13}C,^{15}N-YM	Follow recipe for ^{15}N-YM and replace unlabeled glucose by 2.5 g/L of $^{13}C_6$-glucose
^{2}H,^{15}N-YM and ^{2}H,^{15}N,^{13}C-YM	Prepare ^{15}N-YM or ^{15}N,^{13}C-YM in 99.8 % D_2O. Possibly higher levels of deuteration may be achieved using $^{2}H_7$-glucose or $^{2}H_7$,$^{13}C_6$-glucose respectively. Supply all components as powders to avoid dilution of deuterium
Feeding solution for fed-batch phase	1. Prepare 100 mL of feeding solution per 1 L of fermentation volume based on YM without amino acids and without ammonium, but replace 2.5 g/L by 25 g/L glucose. Use isotope-labeled water and glucose according to your labeling strategy
	2. Adjust solution to pH 6
	3. Filter-sterilize and store at 4 °C
Fermentation of *Pichia pastoris*	
Fermentation in H_2O	1. Grow wild-type *Pichia pastoris* strain (X-33) on fresh LB agar plate at 30 °C
	2. Inoculate unlabeled YM preculture (e.g. 10 mL), grow at 30 °C and 200 rpm (shaking flask) over night to OD_{600} 2
	3. Use preculture to inoculate unlabeled YM scale-up culture (e.g. 100 mL) at OD_{600} 0.2
	4. Grow scale-up culture (shaking flask, 30 °C, 200 rpm) to OD_{600} 2. Harvest cells by centrifugation at 5000*g* for 15 min
	5. Resuspend pelleted cells in labeled YM. Inoculate fermenter to yield a starting OD_{600} of 0.2 in fermentation volume (e.g. 1 L)
	6. Run fermenter-controlled batch-phase at 30 °C, agitation speed 500 rpm, 0.5 L/min airflow, pH 6.0 (adjusted by addition of 4 M NaOH). Maintain dissolved oxygen (DO) level above 35 % by controlled stirring
	7. When initial glucose is fully depleted as judged by increase in DO, start fed-batch phase at a feeding rate to counteract the DO increase
	8. Stop fermentation when fed glucose is depleted. Harvest cells by centrifugation at 5000*g* for 30 min
Fermentation in D_2O	1. Replace unlabeled YM in the scale-up culture with 70 % D_2O-YM for D_2O adaptation
	2. Grow scale-up culture to OD_{600} of 2. Harvest cells by centrifugation at 5000*g* for 15 min
	3. Resuspend cell pellet in 100 % D_2O, labeled YM. Inoculate fermenter to yield a starting OD_{600} of 0.2 in fermentation volume (e.g. 1 L)
	4. Follow H_2O fermentation steps 6–8, but use D_2O feeding and pH control solutions
	5. For recycling of D_2O, supernatants can be sterile-filtered, charcoal-treated, and distilled
Yeastolate preparation	
1. Wash yeast cell pellet with ultrapure water and centrifuge (5000*g*, 30 min)	
2. Prepare a 30 % (w/v) yeast slurry in sterile-filtered water and adjust pH to 7.5 using NaOH	
3. Add 0.5 % (w/w) Zymolyase 20T (Amsbio) and incubate slurry at 35 °C in water bath for 6 h maintaining pH at 7.5	
4. Heat inactivate Zymolyase 20T at 65 °C for 20 min	
5. Adjust to pH 6 and add 0.5 % (w/w) papain powder (Carl Roth)	
6. Incubate slurry under continuous shaking for 5 days at 50 °C	
7. Mechanically homogenize lysate (e.g. using tissue grinder)	
8. Spin down cell debris at 20,000*g* for 30 min and collect supernatant	
9. Filter supernatant through 10 kDa MWCO using a stirred filtration cell (Millipore). Collect flow-through	
10. Lyophilize flow-through to yield a powder of yeastolate. Store at 4 °C	

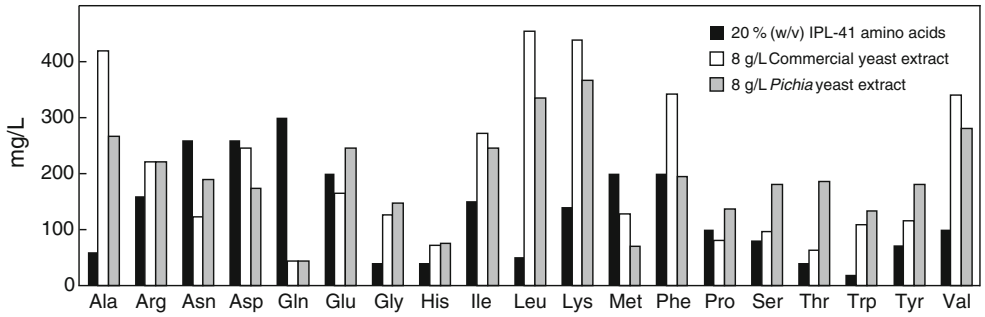


Fig. 1 Amino acid content of different yeast extracts and IPL-41. The free amino acid content of the commercial (Bacto Yeast Extract, Difco, *white*) and *Pichia* yeast extract (*gray*) was quantified by HPLC analysis following pre-column derivatization with phenyl isothiocyanate. Levels of cysteine were not determined by this method. The amino acid content of the yeast extracts is compared to 20 % concentrations of amino acids in the IPL-41 medium (*black*), which was found to be sufficient to promote protein expression in Sf9 insect cells

medium. Interestingly, the overall content of amino acids was comparable in both yeastolates, but differed for individual amino acids. In most cases, the yeastolate amino acid concentrations exceeded those of the 20 % IPL-41 medium. However, glutamine and methionine concentrations were significantly lower in the yeastolates.

The performance of the yeastolate-based insect cell growth media was then assessed by the expression of c-Abl protein in Sf9 cells. Expression in a commercial full medium (Insect Xpress, Lonza) was taken as reference and resulted in yields of 62 mg/L of purified c-Abl (Fig. 2). Following the results of the amino acid analysis, the ΔSF4 medium was initially supplemented by 8 g/L of either commercial or *Pichia* yeastolate without further addition of the low-abundance amino acids glutamine, methionine and cysteine. Under these conditions, expression levels dropped to only 2.4 mg/L purified c-Abl. Hence, the ΔSF4 was further supplemented by 1.0 g/L glutamine and 1.0 g/L methionine (20 % IPL-41). This dramatically increased the yield to 45 mg/L of purified c-Abl, recovering 73 % of the full medium expression. A subsequent test showed that the additional supplementation of methionine was not required, since its omission did not alter the expression level (Fig. 2). Also no further increase was observed upon supplementation of 0.23 g/L cysteine (20 % IPL-41), indicating that addition of cysteine was not necessary. This observation may be explained by previous findings that Sf9 cells can synthesize cysteine from methionine (Doverskog et al. 1998). It has been reported that Sf9 cells are able to synthesize glutamine from ammonium (Drews et al. 2000) and thus glutamine may be substituted by

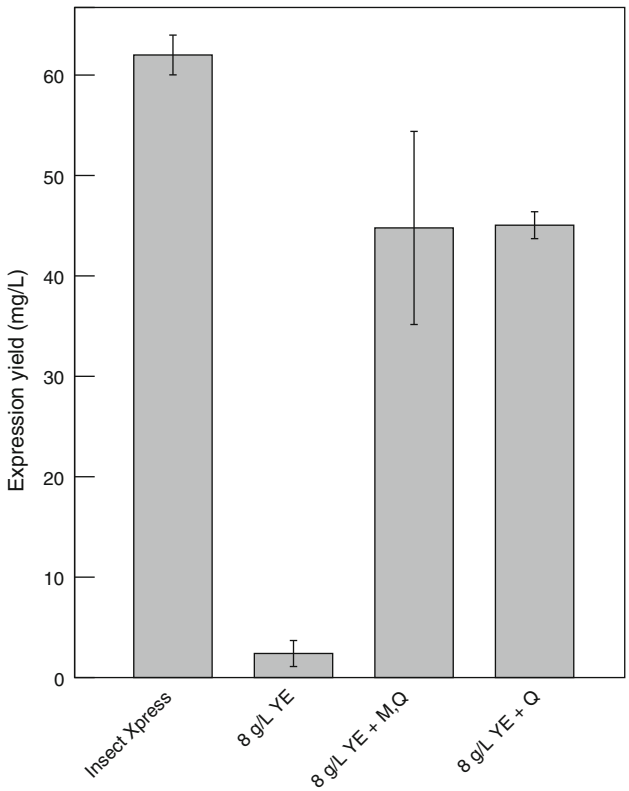


Fig. 2 Expression yield of c-Abl in Sf9 insect cells under various conditions. A reference yield was obtained in the commercial Insect Xpress medium (Lonza). Different supplements to SF4 medium depleted in amino acids and yeastolate (ΔSF4) were tested: 8 g/L of *Pichia* yeastolate (YE), 8 g/L YE + 1 g/L methionine + 1 g/L glutamine (YE + M, Q), and 8 g/l YE + 1 g/L glutamine (YE + Q). Cell cultures in supplemented ΔSF4 were starved for 2 h prior to final medium exchange. All yields were quantified by UV absorption of purified protein at 280 nm

ammonium in Sf9 expression cultures (Meola et al. 2014). However, the substitution by up to 10 mM ammonium resulted in expression levels that were below 10 % of the glutamine-supplemented cultures. Therefore, ΔSF4 supplemented by 8.0 g/L YE and 1.0 g/L glutamine was taken as the optimal medium for all further experiments.

Expression of ^{15}N-labeled c-Abl and trGFPuv

Based on this optimized medium ^{15}N-labeled c-Abl was produced for NMR analysis. Initially, insect cells were grown in unlabeled medium and then changed to the medium containing ^{15}N-yeastolate and $^{15}N_2$-glutamine immediately prior to virus infection. This resulted in yields of 40 mg/L purified protein and an average ^{15}N incorporation of 77 % as determined by mass spectrometry. Since spillover from the initial unlabeled medium was suspected as the source of the low labeling, a starvation step in ΔSF4 without yeastolate and glutamine supplement was introduced 2 h prior to the exchange to ΔSF4 medium with labeled yeastolate and glutamine. The starvation step increased the ^{15}N incorporation to 89 % without reduction of the expression level (Table 3). A similarly high ^{15}N incorporation of 90 % (Table 3) was obtained for the expression of a truncated (Met1-Ile229), optimized form of green fluorescent protein (trGFPuv) (Khan et al. 2003) with a yield of 40 mg/L. Hence, all subsequent expressions were performed with the 2 h starvation step included in the labeling protocol (Table 2).

Figure 3 shows an ^{1}H-^{15}N HSQC spectrum recorded on the ^{15}N-labeled c-Abl protein expressed in insect cells grown on the ^{15}N-yeastolate. The resonances are identical to previously published spectra of c-Abl (Vajpai et al. 2008b). The HSQC resonance intensities were compared quantitatively to an HSQC recorded under identical conditions on a ^{15}N-labeled c-Abl sample prepared from insect cells grown on the commercial, uniformly ^{15}N-labeled BioExpress-2000 (CIL #CGM-2000-N) medium (Strauss et al. 2005; Vajpai et al. 2008b). For the latter, an average ^{15}N labeling efficiency of 91.4 % based on mass spectrometry has been reported. After correction for the respective concentrations, the resonance intensities agreed within ~10 % for all amino acid types (data not shown). This corroborates the very similar ^{15}N incorporation obtained by mass spectrometry and indicates that there is no obvious specific ^{15}N incorporation for certain amino acids.

Table 2 Expression of labeled proteins in insect cells based on yeastolate

Preparation of labeling medium	
Prepare stock solution of yeastolate (e.g. 200 g/L) in sterile ultrapure water. Filter-sterilize and store aliquots at −20 °C	
^{15}N labeling	Supplement dropout medium depleted of amino acids and yeast extract (e.g. ΔSF4, BioConcept) with 8 g/L ^{15}N-yeastolate and 1 g/L $^{15}N_2$-glutamine
^{15}N,^{13}C labeling	Supplement dropout medium depleted of amino acids, yeast extract and all carbon sources (e.g. ΔSF4, BioConcept) with 8 g/L ^{15}N,^{13}C-yeastolate, 5 g/L $^{13}C_6$-glucose and 1 g/L $^{15}N_2$-glutamine
^{2}H,^{15}N,^{13}C labeling	As ^{15}N,^{13}C labeling but use 8 g/L ^{2}H,^{15}N,^{13}C-yeastolate
Adjust pH to 6.2 (±0.2)	
Filter-sterilize medium and store at 4 °C	
Labeled expression in insect cells	
1. Adapt Sf9 or Sf21 cells to growth in serum-free medium (e.g. Insect Xpress, SF4) and optimize expression conditions at 27 °C, shaken at 80 rpm in Erlenmeyer flasks	
2. Scale up cells in mid-log phase of growth to desired expression volume (typically $1.5 \times 10^6 - 2 \times 10^6$ cells/mL)	
3. When the required cell density is reached, harvest cells by centrifugation at 200*g* for 4 min (20 °C)	
4. Resuspend cells in dropout medium of same volume (depleted of amino acids and yeast extract) and starve cells for 2 h (27 °C, 80 rpm)	
5. Determine cell density and viability. Spin down amount of cells required for expression (300*g*, 4 min)	
6. Resuspend cells in labeling medium with half the volume of the final culture	
7. Infect with baculovirus according to optimized conditions (MOI, functional titer). Avoid label dilution by using high-titer stocks	
8. Adjust to final culture volume after 6 h	
9. Grow baculovirus-infected cells for 48–72 h in labeled medium (27 °C, 80 rpm)	
10. Harvest cells at 1000*g* for 10 min and purify recombinant protein	

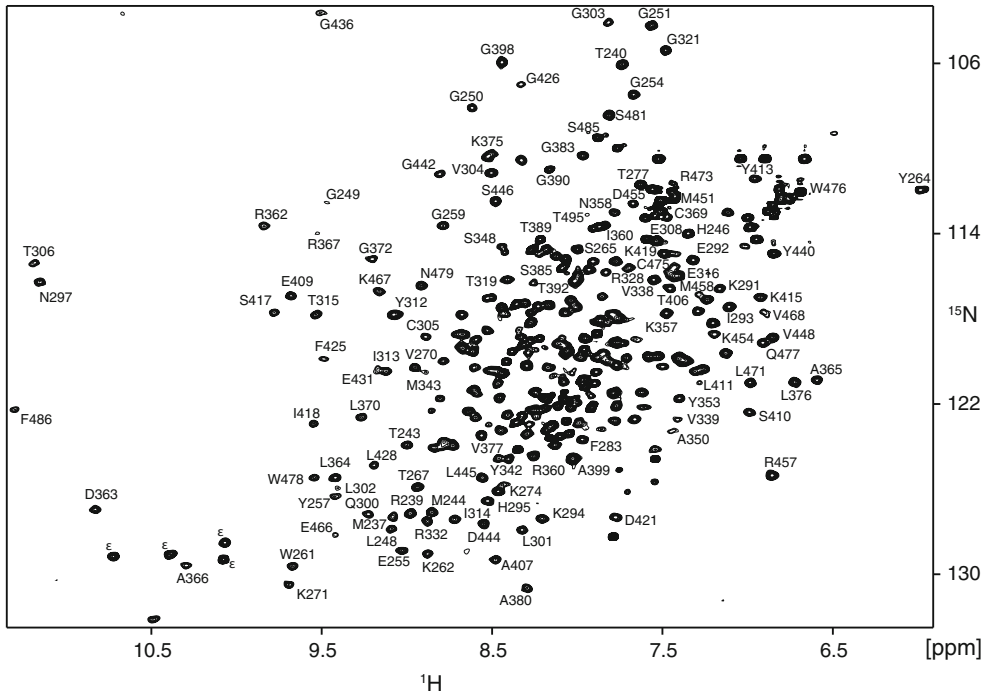


Fig. 3 ^{1}H-^{15}N HSQC spectrum of ^{15}N-labeled c-Abl expressed in Sf9 insect cells. The protein was expressed in Sf9 cells grown on ^{15}N-yeastolate and $^{15}N_2$-glutamine as described in the text. The spectrum was recorded in 2 h on a 200 μM ^{15}N-labeled c-Abl sample

Expression of ^{15}N,^{13}C-labeled trGFPuv and triple resonance NMR

To test its usefulness for canonical triple resonance protein NMR experiments, an ^{1}H,^{15}N,^{13}C-labeled sample of trGFPuv was expressed according to the protocol given in Table 2 using ^{15}N,^{13}C-yeastolate and resulting in a yield of 39 mg/L. The labeling procedure is similar as for the ^{15}N labeling. However, to achieve optimal ^{13}C labeling, the ΔSF4 medium was further depleted of glucose, maltose and sucrose, and 5 g/L of $^{13}C_6$-glucose were supplemented as the only sugar source. Glutamine was only supplemented in

Table 3 Isotope incorporation of proteins expressed using different yeastolate labeling schemes

Protein	Label	Observed mass (Da)	Observed shift (Da)[a]	Expected shift (Da)[b]	Incorporation ratio[c]
c-Abl	Unlabeled	34,892.6[d]	–	–	–
	^{15}N	35,249.2	356.6	402	0.89
	^{15}N,^{13}C	36,558.7	1666.1	1934	0.85[e]
	^{2}H,^{15}N	36,337.4	1444.8	2236	0.61[f]
trGFPuv	Unlabeled	25,585.3	–	–	–
	^{15}N	25,859.6	274.3	305	0.90
	^{15}N,^{13}C	26,778.0	1192.7	1409	0.84[g]

[a] Calculated as the difference between observed mass and unlabeled reference

[b] Theoretical mass shift based on total number of atoms to be labeled

[c] Ratio of the observed by the expected mass shift

[d] The N-terminus of c-Abl kinase domain starts with S*YYHHHHHHDYDIPTTENLYFQGAMDP-c-Abl(S^{229}–S^{500}), * N-acetylation of serine (Strauss et al. 2003)

[e] ^{13}C incorporation assuming 1532 labeled carbon atoms and 0.89 ^{15}N incorporation (356.6 Da)

[f] ^{2}H incorporation assuming 1789 non-exchangable protons and 0.89 ^{15}N incorporation (356.6 Da)

[g] ^{13}C incorporation assuming 1099 labeled carbon atoms and 0.90 ^{15}N incorporation (274.3 Da)

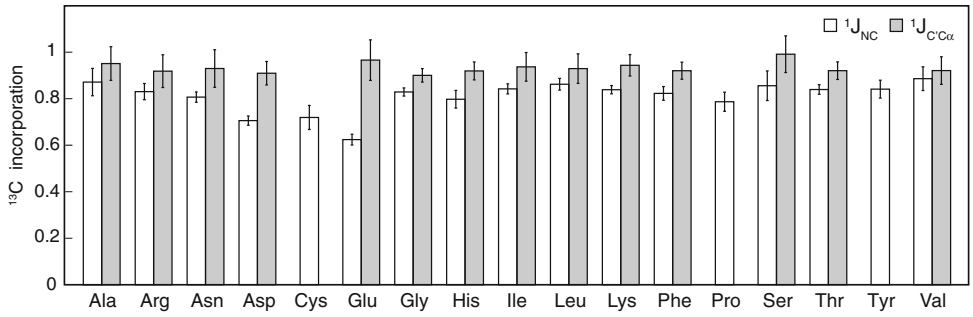


Fig. 4 Specific ^{13}C incorporation in $^{15}N,^{13}C$-GFP analyzed by NMR spectroscopy. $^{15}N,^{13}C$-GFP was expressed in Sf9 cells grown on $^{15}N,^{13}C$-yeastolate and $^{15}N_2$-glutamine. ^{13}C enrichments of the carbonyl nucleus preceding a ^{15}N amide determined via the $^{1}J_{NC'}$ scalar coupling (*white*) and of a C^{α} nucleus next to a $^{13}C'$ carbonyl determined via the $^{1}J_{C'C\alpha}$ scalar coupling (*gray*) are shown as averages and standard errors for the different amino acid types that could be detected. Details are given in Supplementary Figures S1 and S2

^{15}N-labeled form, since addition of the considerably more expensive ^{15}N, ^{13}C-glutamine did not result in a significant increase of ^{13}C labeling as judged by mass spectrometric analysis. An overall ^{13}C incorporation of 84 % was estimated by mass spectrometry, subtracting the previously determined 90 % ^{15}N incorporation from the total mass (Table 3).

The efficiency of ^{13}C incorporation into specific atom positions was further analyzed by NMR spectroscopy. Using modulation by the one-bond $^{1}J_{NC'}$ scalar coupling in the ^{15}N evolution period of a constant-time ^{1}H-^{15}N HSQC (Supplementary Figure S1), the ^{13}C enrichment of the carbonyl nucleus of the amino acid preceding the ^{1}H-^{15}N amide could be determined. Figure 4 shows averages and standard deviations of this enrichment according to the different observed amino acid types. The average ^{13}C incorporation determined by this experiment amounts to 81 %, which is slightly lower than the 84 % determined by mass spectrometry. This reduction may be explained by systematic errors in the NMR determination due the imperfection of the $^{13}C'$ coupling pulse and the faster relaxation of the ^{15}N-{$^{13}C'$} antiphase term, which diminishes the apparent size of the ^{13}C signal (Grzesiek et al. 1993). The amino acid-specific analysis reveals that most amino acids have higher than 80 % (uncorrected NMR value) ^{13}C incorporation at the carbonyl position. However, incorporation is reduced for glutamate (62.4 %), aspartate (71 %), and cysteine (72 %). This is consistent with the reported metabolic conversion of (^{12}C-)glutamine into the latter amino acids in insect cells (Drews et al. 2000).

The 81 % ^{13}C incorporation determined from the one-bond $^{1}J_{NC'}$ scalar coupling corresponds to the probability for finding a $^{13}C'$ nucleus in the amino acid preceding an ^{15}N-labeled amide. However, within a single labeled amino acid the probability for a next neighbor ^{13}C carbon should be higher due to the overall 96 % ^{13}C content of the amino acids coming from the yeastolate (Supplementary Table S1). This was tested by measuring the incorporation of $^{13}C^{\alpha}$ nuclei next to $^{13}C'$ nuclei in a modified HNCO, which was modulated by the one-bond $^{1}J_{C'C\alpha}$ coupling in a constant-time $^{13}C'$ evolution period (Supplementary Figure S2). These intraresidue next neighbor ^{13}C incorporation ratios are also presented in Fig. 4 according to the different amino acid types. No pronounced variations are observed. The average of the intraresidue next neighbor ^{13}C incorporation within a labeled amino acid amounts to 93 % and is clearly considerably higher than the sequential incorporation. Again, the ^{13}C incorporation derived by the NMR experiment presents a lower estimate due to imperfections of the $^{13}C^{\alpha}$ coupling pulse and the faster relaxation of the $^{13}C'$-{$^{13}C^{\alpha}$} antiphase term. The high level of intraresidue next neighbor ^{13}C incorporation agrees with the expected low metabolic conversion of amino acids and has the practical advantage that NMR experiments, which rely on intraresidue heteronuclear connectivities, do not suffer from potentiated losses in sensitivity due to low next neighbor labeling.

The capability to achieve sequential assignments was demonstrated by recording HNCO and HNCA spectra on the $^{15}N,^{13}C$-labeled trGFPuv. Figure 5 shows sequential strip plots from the 3D experiments. Overall, 97 % (89 %) of all expected crosspeaks were observed for the HNCO (HNCA) using an experimental time of 17 h (69 h) on a 230 μM sample of this 26 kDa protein.

Expression of $^{2}H,^{15}N$-labeled c-Abl

To test deuterium labeling, $^{2}H,^{15}N$-labeled c-Abl was expressed in insect cells grown on ΔSF4 medium supplemented by 1 g/L $^{15}N_2$-glutamine and 8 g/L $^{2}H,^{15}N$-labeled

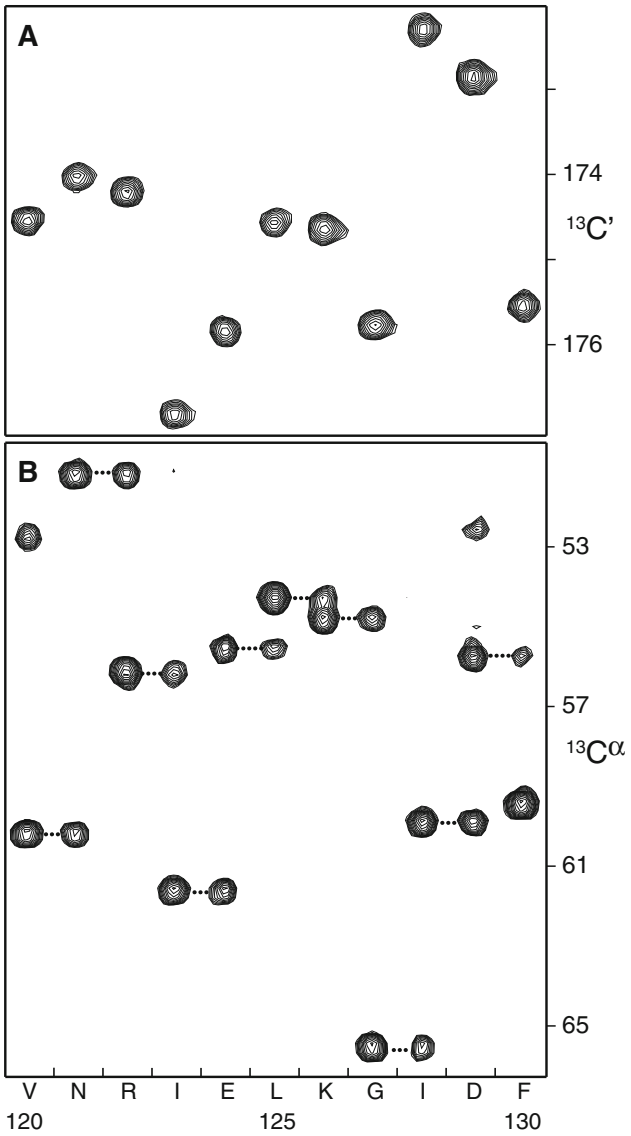


Fig. 5 Triple resonance backbone correlation experiments recorded on ^{15}N,^{13}C-GFP. Strip plots from the 3D data sets of HNCO (**a**) and HNCA (**b**) experiments are displayed for a consecutive stretch of amino acids. ^{15}N,^{13}C-GFP was expressed in Sf9 cells grown on ^{15}N,^{13}C-yeastolate and $^{15}N_2$-glutamine

yeastolate, which had been obtained from a D_2O/$^{15}NH_4Cl$ yeast culture with otherwise protonated medium compounds (Table 1). A yield of 43 mg/L purified ^{2}H,^{15}N-c-Abl was achieved, which is identical to the ^{15}N-labeled or ^{15}N,^{13}C-labeled expression. Mass spectroscopic analysis indicated an average ^{2}H incorporation of 61 % (Table 3), which corresponds to a 19 % drop from the 75 % initial deuteration level of the yeastolate amino acids (Supplementary Table S1).

The benefits of the deuterium labeling were evident in an initial NMR analysis of average amide proton T_2 values, which increased from 10.8 ms in ^{1}H,^{15}N-labeled to 16.2 ms in ^{2}H,^{15}N-labeled c-Abl. An example for the improvement in the spectral sensitivity is given in Fig. 6 for the case of NOESY experiments. Figure 6 shows the amide regions of ^{1}H,^{1}H-NOESY spectra recorded under identical conditions with a mixing time of 230 ms on protonated (Fig. 6a) and deuterated (Fig. 6b) c-Abl. Clearly many more cross peaks become visible for the deuterated case. Furthermore, ^{15}N-edited 3D NOESY spectra were recorded with optimized mixing times of 100 and 230 ms, corresponding to the decay times of the diagonal peaks (selective T_1), for the protonated and deuterated c-Abl, respectively. Obviously, the sensitivity is tremendously increased for the deuterated c-Abl sample (Fig. 6d) and many i, i + 1 and i, i + 2 connectivities are detected, which are not present for protonated c-Abl (Fig. 6c).

Expression of ^{2}H,^{15}N-labeled beta-1 adrenergic receptor

To apply the yeastolate labeling procedure to a highly challenging protein, ^{2}H, ^{15}N-labeled thermostabilized beta-1 adrenergic receptor (β_1AR) (Warne et al. 2008) was expressed in Sf9 cells as described before. A yield of 0.8 mg/L purified, detergent-solubilized protein was achieved, which was comparable to yields from the commercial Insect Xpress (Lonza) medium. The fractional deuteration increased the amide proton T_2 from 3.9 ms observed for the protonated receptor to 5.9 ms for the ^{2}H,^{15}N-labeled receptor. Likewise, the ^{15}N TROSY T_2 increased from 21 to 29 ms upon deuteration. Figure 7 shows the well-resolved TROSY-HSQC recorded for 13 h on the 85 μM ^{2}H,^{15}N-labeled receptor dissolved in a decyl-maltoside detergent micelle (total molecular weight ~100 kDa). About 240 of the 304 expected backbone amide resonances are observable.

Cost analysis

An estimate of the total cost of various isotope labeling schemes in insect cells based on the described YE is provided in Table 4 using current market prices for labeled ingredients and insect cell media. The estimated cost of 1 L insect cell medium for pure ^{15}N labeling amounts to 970 €, which is slightly higher than the estimate of 800 € given for a commercial source of YE-based medium (Meola et al. 2014). However, it should be noted that by the present

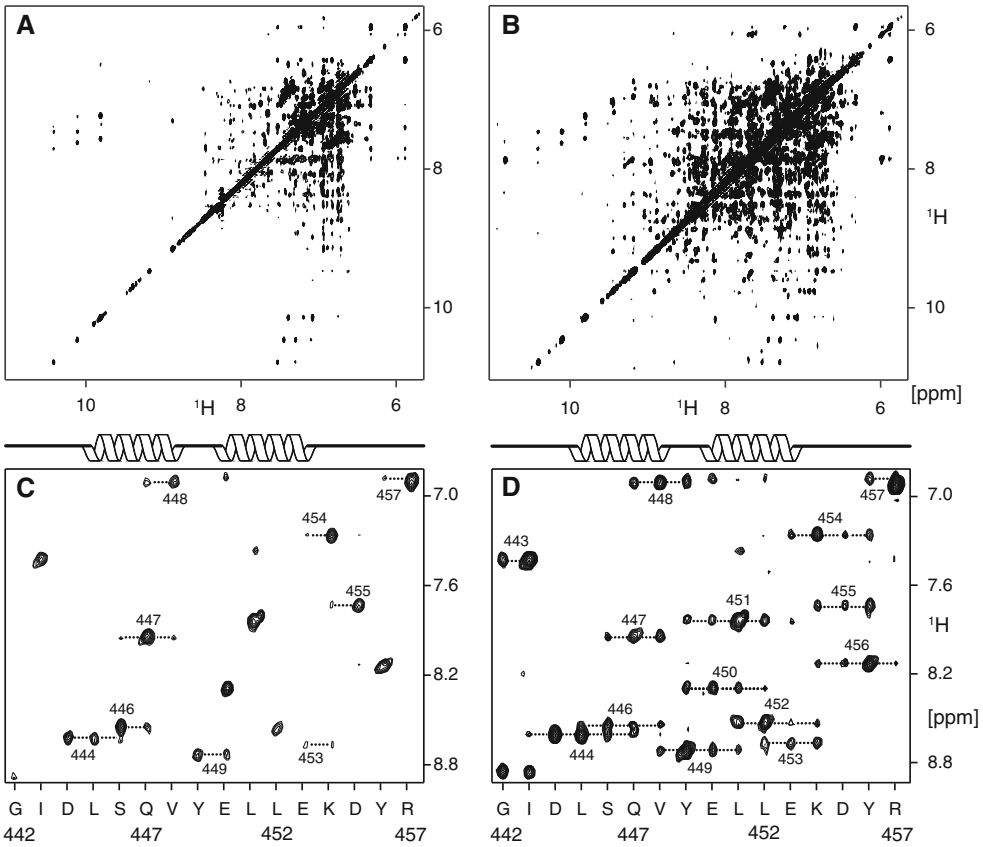


Fig. 6 NOE sensitivity enhancement by deuteration. *Top*: 2D ^{1}H-^{1}H-NOESY spectra recorded under identical conditions on 200 µM samples of ^{15}N-c-Abl (**a**) and ^{2}H,^{15}N-c-Abl (**b**) expressed in Sf9 insect cells (NOE mixing time of 230 ms). *Bottom*: strip plots for a consecutive stretch of amino acids from 3D ^{15}N-edited NOESYs recorded on the same ^{15}N-c-Abl (**c**) and ^{2}H,^{15}N-c-Abl (**d**) samples under identical conditions with the exception of optimized NOE mixing times of 100 ms (^{15}N-c-Abl) and 230 ms (^{2}H,^{15}N-c-Abl)

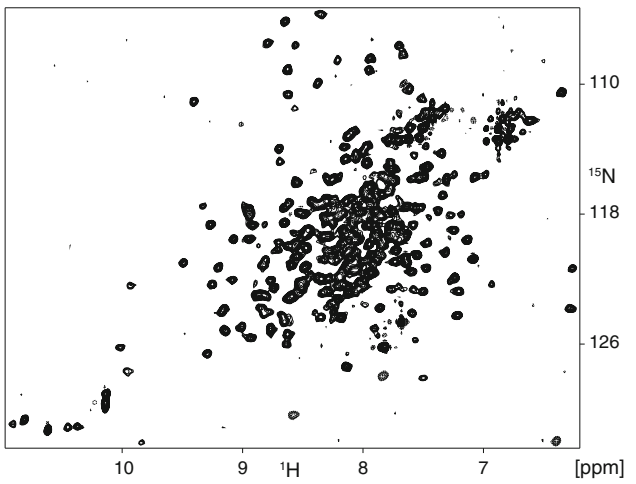


Fig. 7 ^{1}H,^{15}N-TROSY spectrum of ^{2}H,^{15}N-labeled beta-1 adrenergic receptor. ^{2}H,^{15}N-labeled β_1AR was expressed in Sf9 cells grown on ^{2}H,^{15}N-yeastolate and $^{15}N_2$-glutamine as described in the text. The TROSY was recorded in 13 h on 85 µM ^{2}H,^{15}N-labeled β_1AR solubilized in *n*-decyl-β-D-maltopyranoside

method about 4 times higher yields (40 mg/L protein) and 90 % labeling have been achieved without additional costs from a purge phase of insect cell growth on labeled medium. The high expression yield translates into a cost of only 63 € for a typical NMR sample of 0.3 mM Abl kinase domain (MWT 32 kDa) in a 270 µL volume. Furthermore, the largest part of the ^{15}N-labeled medium costs stems from the addition of 1 g/L of ^{15}N-labeled glutamine (800 €). This expense may be reduced to negligible levels by enzymatic synthesis of $^{15}N_2$-glutamine via ^{15}N-glutamate, which can be produced by glutamate dehydrogenase from ^{15}N-labeled ammonium and 2-oxoglutaric acid (Kragl et al. 1993) and subsequently be converted to $^{15}N_2$-glutamine by glutamine synthetase under the addition of ^{15}N-labeled ammonium (Hansen et al. 1992).

Table 4 indicates that compared to ^{15}N labeling, the additional cost for partial deuteration (>60 %) is relatively small (1398 € for 1 L ^{2}H,^{15}N insect cell medium). Since the costs arise from the D_2O used for yeast growth, they may again be reduced considerably by recycling D_2O via distillation. Due to the high amount of glucose required for the yeast growth (27.5 g/L), costs for YE-based insect cell medium increase significantly for ^{13}C labeling and amount to 4424 € for 1 L of ^{15}N,^{13}C insect cell medium. Again the additional cost for the fractional deuteration is relatively minor (4848 € for 1 L ^{2}H,^{15}N,^{13}C insect cell medium). The estimated costs for one sample of triple-labeled Abl kinase domain are still only 314 € under these conditions.

Table 4 Cost estimates for yeastolate-based isotope labeling in insect cells[a]

1 L yeast medium				
Labeling	$^{15}NH_4Cl$[b]	$^{13}C_6$-glucose[c]	D_2O[d]	Total
^{15}N	86	–	–	86
$^{15}N,^{2}H$	86	–	350	436
$^{15}N,^{13}C$	86	2475	–	2561
$^{15}N,^{13}C,^{2}H$	86	2475	350	2911

1 L insect cell medium				
Labeling	Yeastolate[e]	Glucose[f]	Total[g]	c-Abl sample costs[h]
^{15}N	104	–	974	63
$^{15}N,^{2}H$	528	–	1398	91
$^{15}N,^{13}C$	3104	450	4424	287
$^{15}N,^{13}C,^{2}H$	3528	450	4848	314

[a] All costs are given in Euro using conservative estimates of market prices for labeled ingredients and growth media

[b] 4.3 g/L $^{15}NH_4Cl$

[c] 27.5 g/L $^{13}C_6$-glucose

[d] 1 L D_2O

[e] Costs for 8 g isotope-labeled yeastolate based on a yield of 6.6 g per 1 L yeast culture

[f] Addition of 5 g/L $^{13}C_6$-glucose for ^{13}C labeling

[g] Total costs of yeastolate, glucose and 1 g $^{15}N_2$-glutamine (800 EUR) supplements to 1 L of insect cell culture dropout medium (70 EUR, SF4, Bioconcept)

[h] Costs for a 270-µL, 300-µM c-Abl (32 kDa) sample based on an expression yield of 40 mg/L

Conclusion

We have developed a robust protocol for stable isotope labeling of proteins in insect cells based on labeled yeastolate. Our approach overcomes low incorporation rates and low yields described in previous yeastolate-based labeling approaches. This was achieved by a quantitative analysis of the yeastolate amino acid content, supplementation of lacking glutamine, site-specific quantification of isotope incorporation and subsequent optimization. In particular, significant dilution of the isotope labels was prevented by introducing an initial starvation step. In addition to ^{15}N labeling, also efficient $^{15}N,^{13}C$ and partial (>60 %) ^{2}H labeling is demonstrated. Higher ^{2}H incorporation may be achievable by growing yeast on deuterated glucose in addition to D_2O. The described uniform deuteration via the yeastolate may also be combined with the supplementation of specifically labeled amino acids for further gains in resolution and sensitivity (Kofuku et al. 2014).

The preparation of the labeled yeastolate can be readily implemented in a standard, academic structural biology laboratory and does not rely on commercial media of undisclosed content. The costs are moderate for ^{15}N and fractional ^{2}H labeling and may be reduced further by synthesis of labeled glutamine. Labeling by ^{13}C increases the costs, but due to the high yields, still allows production of protonated or partially deuterated $^{15}N,^{13}C$-samples of 'difficult' eukaryotic proteins such as c-Abl at moderate costs of several hundred Euros.

It is hoped that the proposed method will enable detailed studies of eukaryotic proteins such as G-protein-coupled receptors that so far have been inaccessible to NMR due to their requirement of expression in a eukaryotic host. As initial example we provide a well-resolved spectrum of the $^{2}H,^{15}N$-labeled beta-1 adrenergic receptor. Besides their application for insect cells, the described homemade yeast extracts may also serve as a basis for isotope labeling in other higher eukaryotic expression systems such as Chinese hamster ovary and human embryonic kidney cells (Egorova-Zachernyuk et al. 2009).

Materials and methods

Yeast growth and preparation of yeast extract

Detailed protocols are presented in Table 1.

Quantitative amino acid analysis by LC/MS

Crude yeast extract was purified by solid phase extraction using a strong cation-exchange matrix (DSC-SCX SPE, Sigma). Briefly, lyophilized powder was dissolved in 30 mM phosphoric acid solution (pH 2.7) and applied to the conditioned SCX column. The packed column was washed two times with 30 mM phosphoric acid solution. Samples were eluted with 2.5 M NH_4OH containing 10 % acetonitrile and dried in a centrifugal evaporator.

Purified powder was dissolved in 50 % methanol to yield a 10 µg/µL solution for pre-column derivatization with phenyl isothiocyanate (PITC). Two volumes of a freshly prepared coupling solution containing methanol/triethylamine/PITC (7:2:1) were added and the solution was derivatized for 30 min at room temperature protected from light. After derivatization, the mixture was dried under vacuum. Unlabeled standards for each amino acid were prepared accordingly.

The derivatized phenylthiohydantoin (PTH) amino acids were taken up in 20 % acetonitrile for LC/MS analysis. Optimal separation was achieved by HPLC (Agilent) on an Eclipse XDB-C18 reverse phase column (Agilent) at a flow rate of 2.1 mL/min and 50 °C using a gradient elution of 50 mM ammonium acetate at pH 7.2 (Eluent A) against 46:44:10 100 mM ammonium acetate (pH 7.2)/acetonitrile/methanol (Eluent B) (Supplementary Table S2). The PTH

amino acids were detected by optical absorption at 254 and 280 nm, and subsequently identified and quantified for their isotope content on a Bruker microTOF electrospray ionization mass spectrometer. The amount of individual amino acids was determined from the integration of the absorbance peak area and referencing to the unlabeled standard.

Determination of isotope incorporation of expressed proteins by MS

Purified protein samples were desalted by solid phase extraction on C4 reverse phase micro spin columns according to the supplier's protocol (The Nest Group). Isotope incorporation was determined by ESI-TOF mass spectrometry (Bruker microTOF) applying direct infusion of the prepared proteins. Mass spectra were analyzed by maximum entropy deconvolution.

Cell culture and preparation of recombinant baculovirus

In general, handling and maintenance of Sf9 cell culture was based on established protocols (O'Reilly et al. 1994). Permanent cultures were subcultured in mid-log phase and kept in full SF4 medium (Bioconcept) and Insect Xpress medium (Lonza) at 27 °C and shaken at 80 rpm. Recombinant bacmid DNA was generated in DH10Bac *E. coli* cells using the commercial pFastBac vector according to the supplier's procedure (Life Technologies). Baculovirus was generated in adherent cultures as reported by O'Reilly et al. (O'Reilly et al. 1994). Subsequently, high-titer virus stocks for expression were produced by two additional amplification rounds in suspension cultures. In the first amplification round, 50 mL of Sf9 cell culture containing 1×10^6 -cells/mL in serum-free medium were infected with recombinant virus and cultured until cell viability decreased to $\leq$90 %. The supernatant (P1) was collected by centrifugation at 1000*g* and 4 °C for 15 min. For the second amplification round, 100 mL of serum-free Sf9 cell culture containing 1×10^6 cells/mL was infected with P1 and processed accordingly to yield a P2 virus stock.

Functional titration of the P2 virus stock was applied to determine the optimal amount for expression. Briefly, small-scale cultures of Sf9 cells were infected in full SF4 medium at a cell density of 1.5×10^6 cells/mL with P2 virus stock corresponding to functional titers of 1–10 mL per 1 L of cell culture. The yield of purified protein was then quantified for different expression times.

Expression and purification of Abelson kinase domain in insect cells

Samples of Abelson kinase domain (c-Abl) in complex with imatinib were prepared according to the previously published protocol leaving out the TEV cleavage step of the N-terminal polyhistidine tag (Vajpai et al. 2008a). NMR samples were prepared in 20 mM BIS–TRIS (pH 6.5), 150 mM NaCl, 2.5 mM TCEP, 5 % D_2O and 0.02 % NaN_3.

Cloning and purification of truncated trGFPuv

A synthesized gene sequence of the truncated GFP cycle 3 mutant (trGFPuv) based on the previously published construct (Huang et al. 2007) was purchased from a commercial supplier (Genewiz). The sequence was codon-optimized for expression in insect cells. Recombinant baculovirus was prepared by insertion of the trGPFuv gene into the pFastBac vector (Life Technologies). Bacmid DNA was produced in DH10Bac *E. coli* cells according to the standard protocol provided by the supplier.

Functional virus titer and the optimal expression time were determined in small-scale expression tests. trGFPuv was subsequently expressed in SF4-adapted Sf9 insect cells until the cell viability decreased to 90 %. Cells were harvested at 1000*g* for 10 min and lysed by sonication. The protein was purified following the established protocol in *E. coli* (Huang et al. 2007). NMR samples were prepared in 20 mM sodium phosphate (pH 7.4), 5 % D_2O and 0.02 % NaN_3.

Expression and purification of turkey beta-1 adrenergic receptor

Purification of the thermostabilized, turkey beta-1 adrenergic receptor [mutant JM3 (Miller and Tate 2011)] was adapted from the established protocol (Warne et al. 2008). Since the receptor was prepared for solution NMR, solubilization and subsequent purification steps were carried out in *n*-decyl-β-D-maltopyranoside (DM). NMR samples were prepared in 20 mM TRIS/HCl (pH 7.5), 100 mM NaCl, 1 mM alprenolol hydrochloride (racemic), 5 % D_2O, 0.1 % DM and 0.02 % NaN_3.

NMR spectroscopy

NMR experiments were recorded on a Bruker Avance III 900 MHz spectrometer equipped with a triple-resonance cryoprobe. Amide proton T_2 values were determined from

a jump-return spin-echo experiment (Sklenar and Bax 1987). Amide ^{15}N T_2 values were determined from the decay of a ^{15}N spin echo in a TROSY experiment.

All experiments on c-Abl were acquired at a temperature of 25 °C. ^{1}H-^{15}N HSQC experiments on ^{15}N-c-Abl were recorded with 75 (^{15}N) × 1024 (^{1}H) complex points and acquisition times of 30 ms (^{15}N) and 40 ms (^{1}H). 2D ^{1}H,^{1}H-NOESY experiments on ^{15}N-c-Abl and ^{2}H,^{15}N-c-Abl were recorded with 388 (^{1}H) × 2048 (^{1}H) complex points, acquisition times of 18.6 ms (^{1}H) × 80 (^{1}H) ms, and a total experimental time of 6.5 h. The 3D ^{15}N-NOESY-HSQC experiments were recorded with 268 (^{1}H) × 130 (^{15}N) × 2048 (^{1}H) complex points, acquisition times of 16 ms (^{1}H) × 30 ms (^{15}N) × 80 (^{1}H) ms, and a total experimental time of 77 h.

All experiments on trGFPuv were acquired at a temperature of 37 °C. The 3D HNCO experiment on ^{15}N,^{13}C-trGFPuv was recorded with 104 (^{15}N) × 100 (^{13}C) × 2048 (^{1}H) complex points, acquisition times of 27 ms (^{15}N) × 20 ms (^{13}C) × 80 (^{1}H) ms, and a total experimental time of 17 h. The 3D HNCA experiment on ^{15}N,^{13}C-trGFPuv was recorded with 86 (^{15}N) × 82 (^{13}C) × 2048 (^{1}H) complex points, acquisition times of 22 ms (^{15}N) × 9 ms (^{13}C) × 80 (^{1}H) ms, and a total experimental time of 69 h.

The 2D ^{1}H,^{15}N-TROSY experiment on ^{2}H,^{15}N-β_1AR was recorded at 30 °C with 150 (^{15}N) × 1024 (^{1}H) complex points, acquisition times of 30 ms (^{15}N) × 43 (^{1}H) ms, and a total experimental time of 13 h.

Data processing and analysis

All NMR data were processed with the NMRPipe software package (Delaglio et al. 1995). Protein assignments and analysis of peak intensities were carried out using the programs Sparky (Goddard and Kneller 2008) and PIPP (Garrett et al. 1991).

Quantification of ^{13}C incorporation by NMR

Details of the constant-time ^{1}H-^{15}N HSQC and HNCO used for the quantitative assessment of ^{13}C incorporation at the carbonyl and C^{α} position are given in Supplementary Figures S1 and S2.

Acknowledgments We gratefully acknowledge Dr. L. Nisius for initial experiments, Dr. S. Jackson for the gift of the original *E. coli* expression vector of trGFPuv as well as Drs. W. Jahnke, A. Gossert, A. Strauss, L. Skora, G. Schertler, D. Veprintsev, X. Deupi, K. Ballmer-Hofer, T. Maier for very helpful discussions, and I. Hertel for expert help in the preparation of yeast extracts. This work was supported by Swiss National Science Foundation Grants 31-149927 and Sinergia CRSII3-141898.

References

Delaglio F, Grzesiek S, Vuister GW et al (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 6:277–293

Doverskog M, Han L, Häggström L (1998) Cystine/cysteine metabolism in cultured Sf9 cells: influence of cell physiology on biosynthesis, amino acid uptake and growth. Cytotechnology 26:91–102. doi:10.1023/A:1007963003607

Drews M, Doverskog M, Ohman L et al (2000) Pathways of glutamine metabolism in *Spodoptera frugiperda* (Sf9) insect cells: evidence for the presence of the nitrogen assimilation system, and a metabolic switch by ^{1}H/^{15}N NMR. J Biotechnol 78:23–37

Egorova-Zachernyuk TA, Bosman GJCGM, Pistorius AMA, DeGrip WJ (2009) Production of yeastolates for uniform stable isotope labelling in eukaryotic cell culture. Appl Microbiol Biotechnol 84:575–581. doi:10.1007/s00253-009-2063-z

Egorova-Zachernyuk TA, Bosman GJCGM, DeGrip WJ, Shvets VI (2010) Stable isotope labelling of human histamine receptor H1R: prospects for structure-based drug design. Dokl Biochem Biophys 433:164–167. doi:10.1134/S160767291004006X

Garrett D, Powers R, Gronenborn A, Clore G (1991) A common sense approach to peak picking in two-, three-, and four dimensional spectra using automatic computer analysis of contour diagrams. J Magn Reson 95:214–220

Goddard T, Kneller D (2008) SPARKY 3. University of California, San Francisco

Grzesiek S, Vuister G, Bax A (1993) A simple and sensitive experiment for measurement of JCC couplings between backbone carbonyl and methyl carbons in isotopically enriched proteins. J Biomol NMR 3:487–493. doi:10.1007/BF00176014

Hansen AP, Petros AM, Mazar AP et al (1992) A practical method for uniform isotopic labeling of recombinant proteins in mammalian cells. Biochemistry 31:12713–12718. doi:10.1021/bi00166a001

Huang J-R, Craggs TD, Christodoulou J, Jackson SE (2007) Stable intermediate states and high energy barriers in the unfolding of GFP. J Mol Biol 370:356–371. doi:10.1016/j.jmb.2007.04.039

Khan F, Stott K, Jackson S (2003) ^{1}H,^{15}N and ^{13}C backbone assignment of the green fluorescent protein (GFP). J Biomol NMR 26:281–282. doi:10.1023/A:1023817001154

Kofuku Y, Ueda T, Okude J et al (2014) Functional dynamics of deuterated β2 -adrenergic receptor in lipid bilayers revealed by NMR spectroscopy. Angew Chem Int Ed Engl 53:13376–13379. doi:10.1002/anie.201406603

Kragl U, Godde A, Wandrey C et al (1993) Repetitive batch as an efficient method for preparative-scale enzymatic-synthesis of 5-azido-neuraminic acid and N-15-L-glutamic acid. Tetrahedron-Asymmetr 4:1193–1202

Martini AEV, Miller MW, Martini A (1979) Amino acid composition of whole cells of different yeasts. J Agric Food Chem 27:982–984. doi:10.1021/jf60225a040

Meola A, Deville C, Jeffers SA et al (2014) Robust and low cost uniform ^{15}N-labeling of proteins expressed in Drosophila S2 cells and *Spodoptera frugiperda* Sf9 cells for NMR applications. J Struct Biol 188:71–78

Miller JL, Tate CG (2011) Engineering an ultra-thermostable β(1)-adrenoceptor. J Mol Biol 413:628–638. doi:10.1016/j.jmb.2011.08.057

O'Reilly DR, Miller LK, Luckow VA (1994) Baculovirus expression vectors: a laboratory manual. Oxford University Press, Oxford

Rosen MK, Gardner KH, Willis RC et al (1996) Selective methyl group protonation of perdeuterated proteins. J Mol Biol 263:627–636. doi:10.1006/jmbi.1996.0603

Sklenar V, Bax A (1987) Spin-echo water suppression for the generation of pure-phase two-dimensional NMR-spectra. J Magn Reson 74:469–479

Strauss A, Bitsch F, Cutting B et al (2003) Amino-acid-type selective isotope labeling of proteins expressed in Baculovirus-infected insect cells useful for NMR studies. J Biomol NMR 26:367–372

Strauss A, Bitsch F, Fendrich G et al (2005) Efficient uniform isotope labeling of Abl kinase expressed in Baculovirus-infected insect cells. J Biomol NMR 31:343–349. doi:10.1007/s10858-005-2451-3

Vajpai N, Strauss A, Fendrich G et al (2008a) Solution conformations and dynamics of ABL kinase-inhibitor complexes determined by NMR substantiate the different binding modes of imatinib/nilotinib and dasatinib. J Biol Chem 283:18292–18302. doi:10.1074/jbc.M801337200

Vajpai N, Strauss A, Fendrich G et al (2008b) Backbone NMR resonance assignment of the Abelson kinase domain in complex with imatinib. Biomol NMR Assign 2:41–42. doi:10.1007/s12104-008-9079-7

Warne T, Serrano-Vega MJ, Baker JG et al (2008) Structure of a beta1-adrenergic G-protein-coupled receptor. Nature 454:486–491. doi:10.1038/nature07101

Weiss SA, Smith GC, Kalter SS, Vaughn JL (1981) Improved method for the production of insect cell cultures in large volume. In Vitro 17:495–502. doi:10.1007/BF02633510

An economic approach to efficient isotope labeling in insect cells using homemade ^{15}N-, ^{13}C- and ^{2}H-labeled yeast extracts

Supporting Information

Christian Opitz, Shin Isogai and Stephan Grzesiek[*]

Focal Area Structural Biology and Biophysics, Biozentrum, University of Basel, CH-4056 Basel, Switzerland

*Address correspondence to:
Stephan Grzesiek
Focal Area Structural Biology and Biophysics, Biozentrum
University of Basel, CH-4056 Basel, Switzerland
Phone: ++41 61 267 2100
FAX: ++41 61 267 2109
Email: Stephan.Grzesiek@unibas.ch

Table S1: Label incorporation[a] of free amino acids in different yeastolates determined by LC/MS

	Atom number			^{15}N yeastolate		^{13}C yeastolate		^{2}H yeastolate	
	C	*H*[b]	*N*	*Shift (Da)*	*Incorporation*	*Shift (Da)*	*Incorporation*	*Shift (Da)*	*Incorporation*
Ala	3	4	1	0.97	0.97	2.69	0.91	2.54	0.63
Arg	6	7	4	3.68	0.92	5.84	0.98	6.43	0.92
Asn	4	3	2	1.92	0.96	3.82	0.96	2.66	0.89
Asp	4	3	1	0.97	0.97	3.88	0.98	2.44	0.81
Gln	5	5	1	0.91	0.91	4.73	0.95	4.49	0.90
Glu	5	5	2	1.94	0.97	4.76	0.96	4.09	0.82
Gly	2	2	1	0.97	0.97	1.79	0.90	1.65	0.82
His	6	5	3	2.86	0.95	5.91	0.99	3.96	0.79
Ile	6	10	1	0.97	0.97	5.80	0.98	7.40	0.74
Leu	6	10	1	0.97	0.97	5.73	0.96	8.76	0.88
Lys	6	9	2	1.94	0.97	5.55	0.94	7.40	0.82
Met	5	8	1	0.96	0.96	4.80	0.97	5.17	0.65
Phe	9	8	1	0.97	0.97	8.60	0.96	4.71	0.59
Pro	5	7	1	0.97	0.97	4.70	0.95	5.43	0.78
Ser	3	3	1	0.96	0.96	2.82	0.95	1.82	0.61
Thr	4	5	1	0.97	0.97	3.81	0.96	3.21	0.64
Trp	11	8	2	1.89	0.94	10.77	0.99	4.34	0.54
Tyr	9	7	1	0.97	0.97	8.82	0.99	4.19	0.60
Val	5	8	1	0.97	0.97	4.86	0.98	6.62	0.83
Average					0.96		0.96		0.75

[a] Free amino acids were derivatized with phenyl isothiocyanate, separated on a C18 reverse-phase column by HPLC and detected by an ESI-TOF mass spectrometer. Shifts were computed using weighted average values of the observed masses relative to the unlabeled amino acid species. The incorporation was determined as the ratio of the calculated mass shift over the theoretical contribution of the respective atom type.

[b] Non-exchangeable protons.

Table S2: Gradient elution protocol used for amino acid analysis by LC/MS[a]

Step	**Time** (min)	**Eluent A**[b] (%)	**Eluent B**[c] (%)
1	0	100	0
2	6	99	1
3	14	91	9
4	17	65	35
5	20	65	35
6	24	55	45
7	26	0	100
8	29	0	100

[a] The gradient was applied on an Eclipse XDB-C18 column (Agilent) at a flow rate of 2.1 ml/min, with a column length of 150 mm and an inner diameter of 4.6 mm.

[b] 50 mM ammonium acetate at pH 7.2.

[c] 46:44:10 100 mM ammonium acetate (pH 7.2)/acetonitrile/methanol.

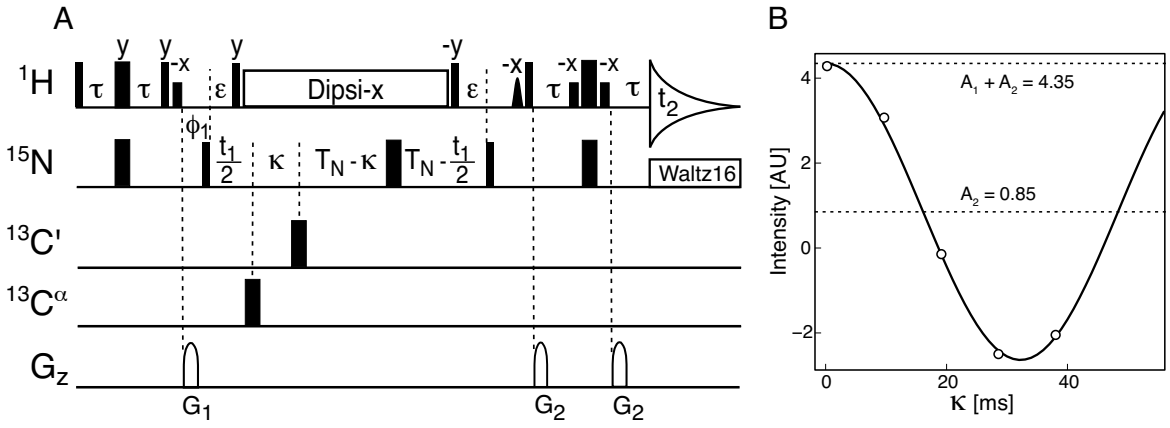


Figure S1: Quantitative NMR analysis of ^{13}C enrichment of the carbonyl nucleus using the modulation by the one-bond $^1J_{NC'}$ scalar coupling in the ^{15}N evolution period of a constant-time ^{1}H-^{15}N HSQC.

A) Pulse scheme of the constant-time $^1J_{NC'}$-modulated ^{1}H-^{15}N HSQC. Narrow and wide pulses have flip angles of 90° and 180°, respectively and are applied with phase x unless specified otherwise. The ^{1}H, ^{15}N, ^{13}C', and $^{13}C^\alpha$ carriers are set to 4.7 (H_2O), 116.5, 177, and 56 ppm, respectively. ^{1}H pulses have an RF field strength of 28 kHz with the exception of the water-selective, low-power, rectangular ~1 ms 90° pulses and the ~2 ms sinc water-selective 90° pulse applied before the reverse ^{15}N-^{1}H INEPT transfer. The ^{15}N pulses have an RF field strength of 6.25 kHz. ^{15}N WALTZ-16 decoupling during acquisition is applied at an RF field strength of 1.5 kHz. The ^{13}C', $^{13}C^\alpha$ 180° decoupling pulses have an RF field strength of 7.0 and 15.7 kHz, respectively. Delay durations are τ = 2.25 ms, ε = 5.4 ms, and T_N = 38 ms. To obtain modulation of crosspeak intensities by the one-bond $^1J_{NC'}$ scalar coupling, the delay κ is incremented in interleaved experiments. For the present experiment, five increments with κ = 0, 9.45, 18.9, 28.35, and 37.8 ms were used. Gradient durations (z-direction, sine bell shaped, 30 G/cm at center): $G_{1,2}$ = 1.0, 0.4 ms. Phases: ϕ_1 = x, -x and ϕ_{rec} = x, -x. Quadrature detection in the ^{15}N dimension is obtained by altering ϕ_1 in the usual States-TPPI manner. Data matrices were recorded as 2048 ($^1H^N$) x 464 (^{15}N) complex points with acquisition times of 80 and 65 ms, respectively. The total experimental times using 16 scans was 15.5 h. Peak amplitudes were obtained by fitting the resonances with the NLINLS program of NMRPipe. The amplitudes were then fitted to the equation $A_1 \cos(2\pi\ ^1J_{NC'}\kappa) + A_2$ by MATLAB (MathWorks) and the relative ^{13}C enrichment of the carbonyl nucleus R was calculated as $R = A_1/(A_1 + A_2)$.

B) Representative fit of the cosine-modulated amplitude for glycine-10 of trGFPuv.

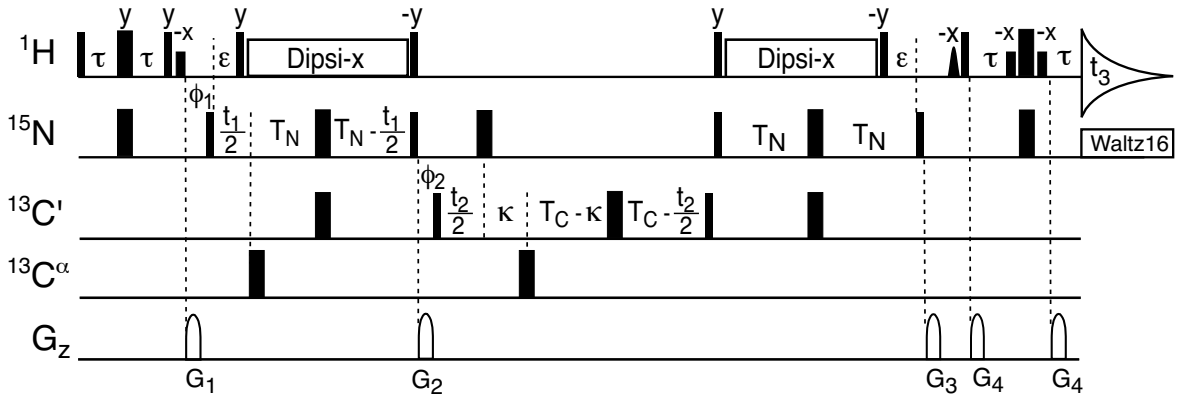


Figure S2: Pulse scheme of the $^1J_{C'C\alpha}$-modulated HNCO used to quantify the ^{13}C enrichment of the C^α nucleus. Narrow and wide pulses have flip angles of 90° and 180°, respectively and are applied with phase x unless specified otherwise. The ^1H, ^{15}N, $^{13}C'$, and $^{13}C^\alpha$ carriers are set to 4.7 (H_2O), 116, 177, and 56 ppm, respectively. ^1H pulses have an RF field strength of 28 kHz with the exception of the water-selective, low-power, rectangular ~1 ms 90° pulses and the ~2 ms sinc water-selective 90° pulse applied before the reverse ^{15}N-^1H INEPT transfer. The ^{15}N pulses have an RF field strength of 6.25 kHz. ^{15}N WALTZ-16 decoupling during acquisition is applied at an RF field strength of 1.5 kHz. The $^{13}C'$, $^{13}C^\alpha$ 180° decoupling pulses have an RF field strength of 7.0 and 15.7 kHz, respectively. Delay durations are τ = 2.25 ms, ε = 5.4 ms, T_N = 13.6 ms, and T_C = 11.2 ms. To obtain modulation of crosspeak intensities by the $^1J_{C\alpha C'}$ scalar coupling, the delay κ is incremented in interleaved experiments. For the present experiment, five increments with κ = 0, 2.2, 4.4, 6.6 and 8.8 ms were used. Gradient durations (z-direction, sine bell shaped, 30 G/cm at center): $G_{1,2,3,4}$ = 1.0, 1.5, 1.0 and 0.4 ms. Phases: ϕ_1 = x, x, -x, -x; ϕ_2 = x, -x and ϕ_{rec} = x, -x, -x, x. Quadrature detection in the ^{15}N and ^{13}C dimensions is obtained by altering ϕ_1 and ϕ_2, respectively, in the usual States-TPPI manner. For the present experiment the ^{15}N dimension was not recorded, and data matrices consisted of 2048 ($^1H^N$) x 202 (^{13}C) complex points with acquisition times of 80 and 22 ms, respectively. The total experimental time using 104 scans was 40.5 h. Peak amplitudes were obtained by fitting the resonances with the NLINLS program of NMRPipe. The amplitudes were then fitted to the equation $A_1 \cos(2\pi\ ^1J_{C\alpha C'} \kappa) + A_2$ by MATLAB (MathWorks) and the selective, intraresidue ^{13}C enrichment of the C^α coupled to the carbonyl nucleus R was calculated as $R = A_1/(A_1 + A_2)$.

2.2 Mass spectrometric analysis of isotope-labeled proteins expressed in insect cells

Introduction

Metabolic scrambling and dilution of the isotope label are key challenges for expression of highly deuterated proteins in insect cells. High levels of deuteration can only be obtained by providing the majority of the amino acid sources in the growth medium in their deuterated form. However, so far the highest reported level of deuteration was only ~76 % based on the use of perdeuterated algal extracts (*39*). A thorough understanding of the metabolic pathways causing the scrambling of the deuterium label may help to improve deuteration in insect cells.

Whereas incorporation of deuterium has been studied extensively in *E. coli* (*99*) and in the associated cell-free expression system (*100*, *101*), it remains poorly characterized in insect cells and in fact all higher eukaryotic expression systems. A very recent investigation of deuterium incorporation in insect cells has been carried out for the partially deuterated β_2-adrenergic receptor studied by NMR (*41*), which was grown on deuterated algal extracts and deuterated amino acids supplemented at specific time points during expression and analyzed by NMR. The observed deuteration levels were not uniform and varied between not significant and 90 %. A strategy to achieve higher levels of deuteration remains unclear.

This section describes a mass spectrometric approach to study the isotope incorporation of uniformly labeled proteins expressed in insect cells. The method allows rapid and simultaneous analysis of ^{2}H and ^{15}N incorporation resolved for individual amino acids on sub-microgram sample quantities at any time point during expression. An overall dilution of the isotope label was observed for all analyzed amino acids, which was more pronounced in the case of deuteration. It is very likely that a significant cellular pool of unlabeled amino acids causes this dilution. The analysis further indicated that deuteration was significantly lower for several non-essential amino acids, such as alanine, glutamate, glycine, and aspartate. Finally, possible optimizations of deuterium labeling in insect cells are discussed.

Remark on author contribution

The work of this section was carried out together with Jethro Hemmann (temporarily member of the lab of Professor Stephan Grzesiek fulfilling compulsory paid community service, now University of Zurich) who performed expression, purification, and mass spectrometric analysis of ^{2}H-labeled GFP.

Results and Discussion

Amino acid-resolved isotope incorporation of $^{2}H,^{15}N$-labeled Abelson kinase

Mass spectrometric analysis of single amino acids was established using $^{2}H,^{15}N$-labeled Abelson kinase (cAbl) expressed in Sf9 insect cells grown on 8 g/L $^{2}H,^{15}N$-labeled yeast extract (YE), prepared by fermentation of yeast in D_2O using protonated glucose, and 1 g/L $^{15}N_2$-glutamine. $^{2}H,^{15}N$-labeled cAbl was chosen as a reference based on the previous analysis of isotope incorporation on the whole protein level by mass spectrometry and NMR (see Chapter 2.1). The purified protein was hydrolyzed in 6 M hydrochloric acid to obtain free amino acids, which were further modified by a pre-column derivatization step to facilitate

LC/MS analysis (see Materials and Methods). Analysis of the isotope incorporation was not possible for cysteine and tryptophan, as acid hydrolysis results in oxidation of the two amino acids and impedes their recovery (*34*).

Furthermore, glutamine and asparagine are converted to glutamate and aspartate, respectively (*32*, *34*). The isomeric amino acids leucine and isoleucine cannot be distinguished based on their atomic mass. Consequently, isotope incorporation of D, E, and I reflects a pairwise weighted average with N, Q, and L corresponding to their individual abundance in cAbl.

The average incorporation of ^{15}N was computed as the combined weighted average of the single amino acids obtained by hydrolysis of $^{2}H,^{15}N$-labeled cAbl (Figure 2.1A, red bars) yielding an average incorporation of 89.4 %. This agrees very well with an incorporation level of 89 %, which has been determined on the whole protein using mass spectrometry (Chapter 2.1). Remarkably, the ^{15}N incorporation was homogenous for most of the amino acids. However, methionine displayed a significantly lower incorporation level (75.6 %). Yeast extracts have low abundance in methionine, which is not synthesized *de novo* in insect cells (*38*). As supplementation of methionine was not required for high-level expression in the yeast extract-based growth medium (see Chapter 2.1), this hints at a significant intracellular amount of methionine. Consequently, unlabeled methionine may be incorporated into the protein and may reduce the overall ^{15}N incorporation.

In contrast, incorporation of ^{2}H was strongly dependent on the amino acid type (Figure 2.1A, grey bars). The average level of deuteration based on the combined weighted average of the single amino acids was only 49.3 %. In the case of alanine, aspartate, and glutamate ^{2}H incorporation was significantly reduced to 4 %, 14 %, and 3.5 %, respectively. These non-essential amino acids were essentially protonated, which is likely the result of metabolic scrambling and may thus be rationalized by the amino acid metabolism of insect cells (Chapter 1.3). Based on the amino acid composition of cAbl (see Chapter 2.1), an average ^{2}H incorporation of 55 % is expected using the weighted average ^{2}H incorporation of individual amino acids (Figure 2.1A, grey bars). This corresponds to a decrease in the deuteration level by 20 % in insects cells with respect to the original 75 % of the nutrient yeast extract (Chapter 2.1, Table S1). The previous analysis of the isotope incorporation on the whole protein had indicated a slightly higher average ^{2}H incorporation of ~60 % for $^{2}H,^{15}N$-labeled cAbl (see Chapter 2.1), which may be due to the oversimplifying assumption that the ^{15}N incorporation was identical for different preparations of isotope-labeled cAbl.

To further understand the apparent difference in the isotope incorporation related to analytical approach, the distribution of ^{2}H and ^{15}N isotopologues was analyzed on a single amino acid level. The ^{15}N incorporation was similar for most of the amino acids with the exception of methionine (Figure 2.1B, red bars). In contrast, analysis of the ^{2}H incorporation (Figure 2.1B, grey bars) revealed significant fractions of completely protonated isotopologues for all detected amino acids. As only ^{15}N-labeled sources of amino and amide nitrogen including $^{15}N_2$-glutamine have been supplemented to the growth medium, this observation strongly suggests biosynthesis of non-essential amino acids from fully protonated precursors and ^{15}N-labeled amino acids (Chapter 1.3). Whereas the sources of unlabeled nitrogen are limited in the growth medium, back-protonation may be caused by several components such as the solvent, α-keto acids, and carbohydrates.

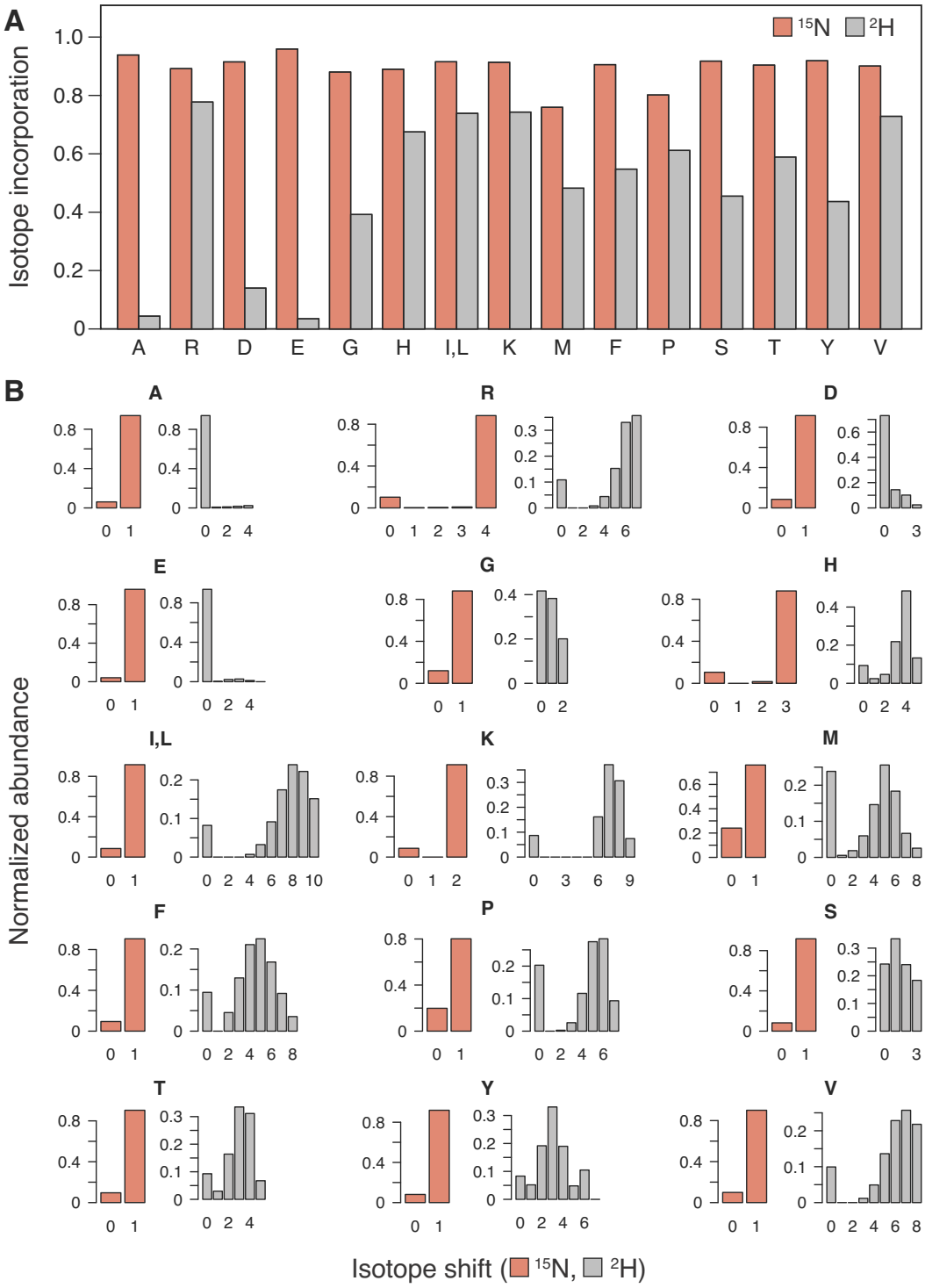


Figure 2.1 Amino acid-resolved isotope incorporation of $^{2}H,^{15}N$-labeled cAbl analyzed by mass spectrometry. **A** Weighted average incorporation of ^{15}N (red) and ^{2}H (grey) was determined for single amino acids obtained by acid hydrolysis of $^{2}H,^{15}N$-labeled cAbl expressed in Sf9 insect cells grown on $^{2}H,^{15}N$-labeled yeast extract (see Materials and Methods). **B** Isotopologue distributions of individual amino acids. The isotope shift indicates the number of labeled nitrogen atoms (red) or aliphatic deuterons (grey) in the amino acid including the fully unlabeled species (isotope shift of 0).

Compared to the ^{2}H,^{15}N-labeled yeast extract (Chapters 2.1, Table S1), an average reduction of deuteration level by ~9 % was apparent for the essential amino acids F, I, K, L, M, T, and V (Figure 2.1B, grey bars). As metabolic scrambling is limited to non-essential amino acids (Chapter 1.3), such a reduction in the overall deuteration level was likely caused by a completely unlabeled fraction of essential amino acids. In the case of ^{15}N labeling, the overall incorporation was reduced to 90 % (Figure 2.1A, red bars). In the absence of further metabolic scrambling, a similar decrease would be expected for the deuteration. However, the observed 55 % deuteration is significantly lower than this value. To further investigate this effect, a mass spectrometric analysis of the initial level of deuteration in the deuterated yeast extracts has to be performed in further experiments.

Analysis of ^{2}H incorporation for growth of insect cells on perdeuterated algal extracts

The ^{2}H incorporation depends on the initial level of deuteration in the labeled extract, which was so far prepared by autolysis of yeast grown on D_2O but not using any other deuterated ingredients (see Chapter 2.1). As a result, an average ^{2}H incorporation of 75 % was obtained for such an extract. Thus, it may be possible to increase the deuteration in insect cells using perdeuterated yeast extracts, which requires growth of yeast in fully deuterated growth media.

Alternatively, commercial perdeuterated algal extracts (D-ISOGRO, Sigma-Aldrich) can be employed as a source of highly deuterated amino acids in insect cells (*39*). To test whether deuteration in insect cells may be further increased using perdeuterated algal extracts, isotope incorporation was first analyzed on the free amino acids in ^{2}H-labeled algal extracts using mass spectrometry. The overall deuteration in these extracts was computed using the weighted average for all detected amino acids (Figure 2.2A, blue bars) and amounts to 94.7 %. This is slightly lower compared to the expected deuteration level of 98.5 % according to the supplier's information. In most of the cases, the average deuteration of the single amino acids exceeded 90 %. However, deuterium labeling of histidine and tyrosine was significantly lower, amounting to only 83.6 % and 73.7 %, respectively.

The analysis of isotopologue pattern for each amino acid in the perdeuterated algal extract (Figure 2.2B, blue bars) indicated very high levels of deuterium incorporation corresponding to the expected shift based on the non-exchangeable aliphatic sites of the individual amino acid. Compared to the isotopologue patterns observed for only partially deuterated amino acids obtained by hydrolysis of ^{2}H,^{15}N-labeled cAbl, low levels of residual protonation were observed, which are caused by growth of algal cells on D_2O with a typical deuteration level of 99.8 %. This slightly increases the residual protonation of amino acids with long aliphatic or complex side chains such as isoleucine, arginine or phenylalanine according to a binomial distribution. Furthermore, only partial deuteration of histidine (Figure 2.2B, H) and tyrosine (Figure 2.2B, Y) was observed in the perdeuterated algal extracts. It is very unlikely that only two amino acids are partially deuterated in an otherwise perdeuterated algal extract prepared by growth of algal cells in D_2O for several generations. Therefore, back-protonation of histidine and tyrosine has possibly occurred during acid hydrolysis of the protein or during the subsequent derivatization step. To test whether this effect is caused by hydrolysis of the protein, direct analysis can be performed on $^{2}H_5$-histidine and $^{2}H_7$-tyrosine, which are available from commercial suppliers. It is expected that the same effect was present in the analysis of H and Y obtained by acid hydrolysis of ^{2}H,^{15}N-labeled cAbl.

In the next step, incorporation of deuterium was investigated using the green fluorescent protein (GFP) expressed in Sf9 insect cells using 8 g/L ^{2}H-labeled algal extract and 1 g/L ^{1}H-glutamine (see Chapter 4). Again, the purified protein was subjected to acid hydrolysis and pre-column derivatization prior to LC/MS analysis. Consequently, similar considerations apply for the recovery and analysis of certain amino acids (C, D, E, I, L, and W) as described above. Most notably, the overall deuteration level was significantly reduced to 53.4 %, which is only slightly higher than the level obtained for ^{2}H,^{15}N-labeled cAbl (49.3 %). The weighted average incorporation of individual amino acids also indicated differential back-protonation (Figure 2.2A, green bars) and the overall deuteration pattern of the amino acids obtained by hydrolysis of ^{2}H-labeled GFP resembles the one observed for ^{2}H,^{15}N-labeled cAbl (Figure 2.1A, grey bars). This hints at metabolic scrambling irrespective of the expressed protein. Again, differential incorporation of deuterium was observed based on the amino acid type. In particular, a substantial reduction in the deuteration level to 4 %, 0.4 %, and 1.5 %, was observed for the non-essential amino acids alanine, aspartate, and glutamate, respectively (Figure 2.2A, green bars). For expression of ^{2}H-labeled GFP, cultures were grown on dropout media containing 5 g/L glucose and 1 g/L glutamine, which provide protonated carbon scaffolds for biosynthesis of A, D, and E (Chapter 1.3). Therefore, metabolic scrambling may explain the significant reduction of the ^{2}H incorporation for these amino acids. In the original protocol (Chapter 2.1), the use of deuterated glucose has been explored, but no improvement was obtained. This suggests that glutamine is the major source of the protonated carbon scaffold and deuteration of glutamine is likely to increase deuteration in insect cells.

The distribution of the ^{2}H isotopologues was analyzed for each amino acid obtained by hydrolysis of ^{2}H-labeled GFP (Figure 2.2B, green bars) indicating that in all amino acids one fraction was completely protonated. Although back-protonation of single aliphatic sites was observed, its contribution to the dilution of isotope label was insignificant. Whereas in most cases only the fully protonated isotopologue was detected, significant changes were observed for A, D, E, G, and S (Figure 2.2B, green bars). As discussed above, these non-essential amino acids are subjected to significant metabolic scrambling involving important anaplerotic and housekeeping reactions in insect cells (Chapter 1.3).

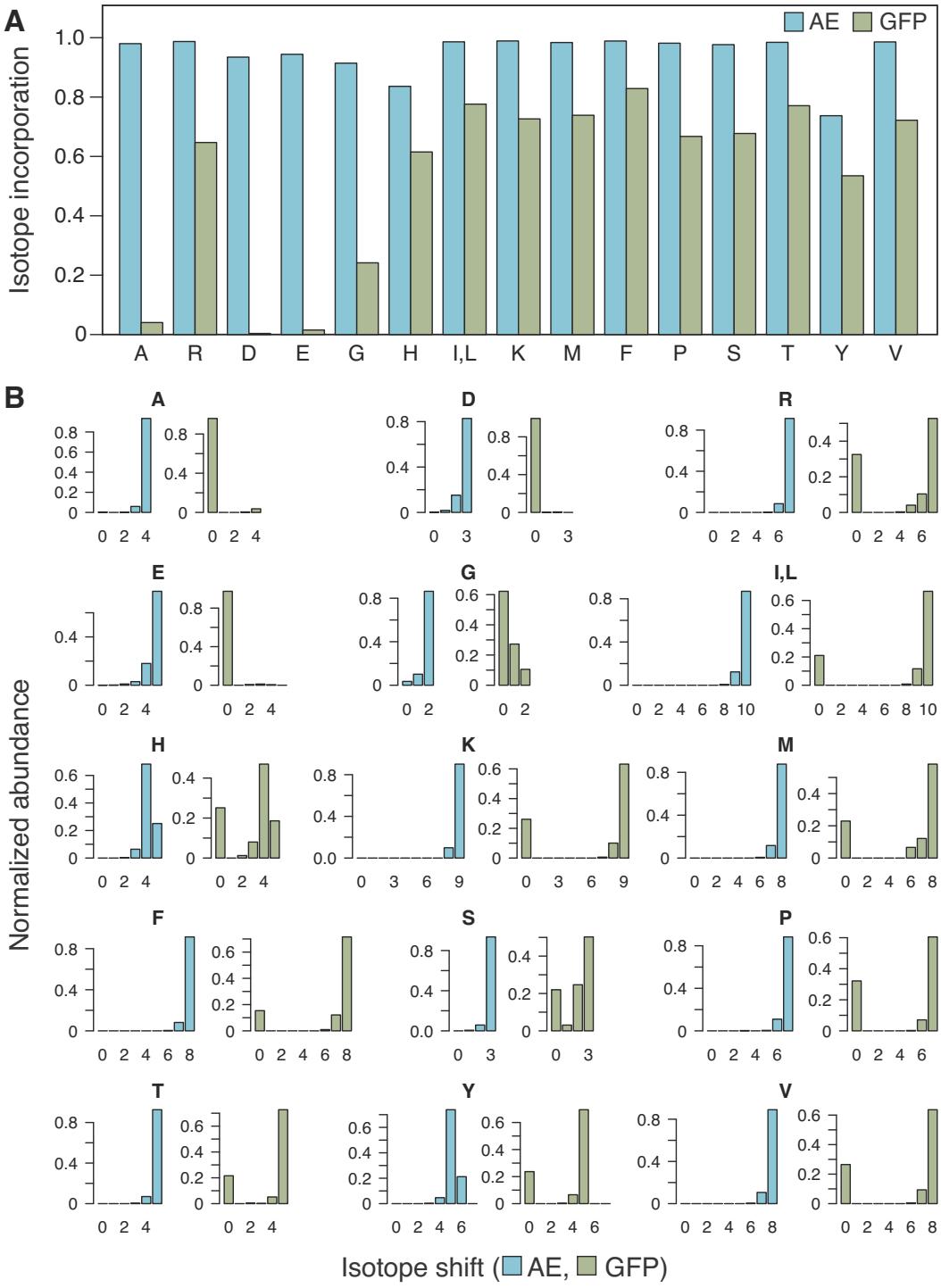


Figure 2.2 Amino acid-resolved incorporation of deuterium analyzed by mass spectrometry. **A** Weighted average deuteration of single amino acids determined by direct analysis of the free amino acids in ^{2}H-labeled algal extract (blue) and by acid hydrolysis of ^{2}H-labeled GFP (green) expressed in Sf9 insect cells grown on ^{2}H-labeled algal extract (see Materials and Methods). **B** Isotopologue distribution for individual amino acids in ^{2}H-labeled algal extract (blue) and amino acids obtained by acid hydrolysis of ^{2}H-labeled GFP (green). The isotope shift indicates the number of aliphatic deuterons in the amino acid including the fully protonated species (isotope shift of 0).

Remarkably, the results obtained for perdeuterated algal extracts strongly correlate with the mass spectrometric analysis of the amino acids obtained by hydrolysis of $^{2}H,^{15}N$-labeled cAbl. Moreover, the average incorporation of ^{2}H was comparable for perdeuterated algal and partially deuterated yeast extracts. At first, this is counterintuitive, as the initial deuteration level is significantly higher in the perdeuterated algal extract (~95 %). However, analysis of the isotopologue distribution indicated that a major fraction of fully protonated amino acids was incorporated into the labeled proteins. In the case of ^{2}H-labeled GFP, this protonated fraction was increased compared to $^{2}H,^{15}N$-labeled cAbl. Furthermore, reduction of the ^{2}H incorporation was also observed for essential amino acids, albeit less pronounced. As insect cells cannot synthesize such amino acids, this again hints at a cellular pool comprising all amino acids in their unlabeled form. Hence, the comparable level of deuteration may be explained by two reasons: (i) a different size of the amino acid pool due to metabolic variations of the insect cell and (ii) the relative amount of amino acids provided to the growth medium by supplementation of the cellular extracts.

In a recent protocol, high deuteration levels of ~76 % were reported using isotope-labeled algal extracts (*39*). However, the protocol also uses ^{15}N-labeled ammonium salt to avoid supplementation of labeled glutamine. When glutamine is not present in the growth medium, insect cells synthesize glutamine from glutamate and ammonium by activity of glutamine synthetase (*35*). Consequently, deuterated glutamine can be formed using deuterated glutamate, which is highly abundant in the commercial algal extracts (Chapter 4).

Conclusions and perspectives

A mass spectrometric approach was established to study the isotope incorporation of isotope-labeled proteins expressed in insect cells. The method enables simultaneous analysis of different isotopes such as ^{2}H and ^{15}N at the single amino acid level. It was shown that the results obtained by mass spectrometric analysis were in good agreement with the NMR data described in Chapter 2.1. The data hints at a significant dilution of the isotope label due to a cellular pool of unlabeled amino acids, which caused an overall decrease of the labeling. Remarkably, incorporation of ^{15}N was homogenous for all amino acids, whereas the level of deuteration was strongly dependent on the amino acid type. The cellular pool of amino acids may be further studied in insect cell cultures, which are continuously grown on dropout media supplemented with isotope-labeled yeast extract. Isotope labeling for individual amino acids may then be determined within the cell in comparison to the medium at several time points.

Substantial back-protonation was predominantly observed for non-essential amino acids, most notably for alanine, glutamate, and aspartate. This may be related to the use of protonated glucose or glutamine, as their concentrations exceed those of the deuterated amino acids provided by supplementation of the cellular extracts. It is hoped that further insights into metabolic scrambling and label dilution can be obtained by mass spectrometric analysis, which may then be used to guide optimization of the existing labeling protocols (*39*, *98*) and increase the level of deuteration.

In the next step, uniform labeling by ^{13}C will be analyzed using ^{13}C-labeled yeast and algal extracts (Chapter 2.1 and Chapter 4). This may be further combined with the use of ^{13}C-glucose to estimate the metabolic scrambling of the glucose carbon scaffold. It is well

known that ^{13}C is scrambled into several non-essential amino acids. However, so far the analysis has only been performed on the whole protein using mass spectrometry (*13*).

The method can also be applied to investigate isotope incorporation for labeling schemes using a single labeled amino acid in an otherwise unlabeled growth medium. In particular, metabolic scrambling of glutamine has to be explored using ^{13}C- and possibly ^{2}H-labeled glutamine in further experiments. Selectively deuterated glutamine can be obtained by enzymatic synthesis as described in Chapter 5.

Finally, mass spectrometric analysis may be used to identify differences in the label incorporation between permanent cultures and cultures infected with the baculovirus. Recent metabolic analysis hints at a change of amino acid utilization upon infection (*96*). As expression is driven by very late viral promoters and occurs only after several hours post infection, additional exchange steps in labeled growth medium may be performed to overcome some of the observed limitations.

Materials and Methods

Uniform isotope labeling in insect cells

Detailed protocols are presented in Chapter 2.1.

Purification of isotope-labeled cAbl and trGFPuv

Detailed protocols are presented in Chapter 2.1

Acid hydrolysis of proteins

Free amino acids were obtained by acid hydrolysis of 1-10 µg of purified protein using the analysis service of the Functional Genomics Center Zurich (FGCZ, Zurich).

Pre-column derivatization of free amino acids

Detailed protocols are presented in Chapter 2.1

LC-MS/MS analysis of free amino acids

Following derivatization, 0.5 µg of each sample was dissolved in solvent A (2 % acetonitrile, 0.15 % formic acid) and subjected to LC-MS analysis using a dual pressure LTQ-Orbitrap Elite mass spectrometer connected to an electrospray ion source (Thermo Fisher Scientific) as described recently (*102*) including minor modifications. Peptide separation was performed on an EASY nLC-1000 system (Thermo Fisher Scientific) equipped with a RP-HPLC column (75 µm × 13 cm) packed in-house with 1.9 µm ReproSil-Pur 120 C18–AQ resin (Dr. Maisch GmbH) applying a linear gradient from 95 % solvent A and 5 % solvent B (98 % acetonitrile, 0.15 % formic acid) to 50 % solvent B over 15 min at a flow rate of 0.25 µl min^{-1}. The data acquisition mode was set to obtain one high resolution MS scan in the FT part of the mass spectrometer at a resolution of 240,000 full width at half-maximum (at m/z 400) followed by MS/MS scans in the linear ion trap of the 10 most intense ions using a collision energy of 25 %. Charged state screening was applied to exclude unassigned charged ions using a dynamic exclusion duration of 15 s. Ion accumulation times were set to 300 ms (MS) and 50 ms (MS/MS), respectively.

3 Thermostabilization and isotope labeling of the β_1-adrenergic receptor to study dynamics and ligand interactions by NMR

Organization of the chapter

The first section (Chapter 3.1) describes the initial design of a thermostabilized mutant of the turkey β_1-adrenergic receptor (TS-β_1AR) for expression and isotope labeling in insect cells and its characterization. Thermostabilization was optimized as a combination of previously established mutants for solution NMR studies.

The following section (Chapter 3.2) is a reprint of the original publication Isogai, Deupi, Opitz et al. (*103*), which describes the use of TS-β_1AR in combination with ^{15}N-valine labeling to study the response of the receptor upon binding of six distinct ligands at high resolution. Further, the influence of thermostabilization was investigated by reverting mutations towards the wild type sequence based on the fully stabilized receptor. Finally, the last section (Chapter 3.3) reports an approach to backbone assignment using high-quality ^{15}N-edited NOESY spectra obtained on a ^{2}H,^{15}N-labeled TS-β_1AR expressed in insect cells, which were grown on ^{2}H,^{15}N-labeled yeast extracts as described in Chapter 2.1.

3.1 Rational design of a turkey β_1-adrenergic receptor mutant for NMR studies

Introduction

G-protein coupled receptors (GPCRs) are a physiologically important class of integral membrane proteins, which transduce extracellular signals to the intracellular environment. Numerous efforts have been made to study structure and function of GPCRs. However, the low stability of wild type receptors often precludes structural studies of their native form. Hence, virtually all crystallographic approaches to structure determination of GPCRs rely on extensive engineering to obtain stabilized receptors (*104*). Whereas ligand binding is still observed in these studies and used to further stabilize GPCRs for crystallization, active signaling of the receptor may be strongly reduced such as for the stabilized β_1-adrenergic receptor (β_1AR) (*105*, *106*).

Recent technical progress in NMR spectroscopy has allowed to study ligand binding to eukaryotic GPCRs (*29*, *41*, *107*) and more impressively to determine the structure of the chemokine receptor CXCR1 (*108*). However, so far the contribution of NMR to characterize dynamics of eukaryotic receptors has been limited to very selective information obtained by labeling of amino acid side-chains (*41*, *107*, *109*). This is largely due to the difficulties of isotope labeling in eukaryotic expression systems used to obtain an active form of the receptor. Moreover, solubilization of GPCRs relies on membrane mimics increasing the overall size of these complexes to the practical limit of solution state NMR.

This chapter describes an approach to thermostabilization of the β_1-adrenergic receptor (β_1AR) for NMR spectroscopic analysis in solution. Rational design of the construct followed previous approaches utilized in crystallographic studies (*103*, *105*, *110*). Expression of the mutant receptor (TS-β_1AR) was established in insect cells. Moreover, the stability of the TS-β_1AR was analyzed by a microscale thermal shift assay indicating a melting temperature of 57 °C, which corresponds to an increase by 31 °C compared to the wild type receptor (*104*, *110*) and is highly beneficial for NMR studies.

Results and Discussion

Thermostabilization of β_1AR for NMR studies

Rational design of a thermostabilized β_1AR (TS-β_1AR) for NMR studies was based on mutations used in two recent crystallographic studies (Figure 3.1) (*105*, *110*). The stabilized *turkey* β_1AR44-m23 mutant contains eight thermostabilizing mutations, namely R68S, M90V, C116L, Y227A, A282L, F327A, F338, and C358A (Figure 3.1, orange residues). The C116L point mutation was reported to increase the yield of purified receptor, whereas C358A removes the C-terminal palmitoylation site to aid in crystallization. All remaining mutations were shown to increase the thermostability of the receptor to a different extent (*111*), which eventually increased the melting temperature by 21 °C over the native β_1AR. Moreover, none of these mutations is directed at amino acids essential to ligand binding (Figure 3.1, purple residues) but significantly decreased the affinity for agonists such as isoprenaline (*111*). In contrast, this effect was less pronounced for binding of antagonists, which infers stabilization of the receptor in the inactive state (*106*, *111*). β_1AR44-m23 is further truncated at both termini (Δ3-32 and Δ369-483) and the intracellular loop ICL3 (Δ243-270 and Δ247-278). Whereas truncation of the N-terminus prevents proteolysis, truncation of ICL3 leaves the loop still flexible and allowed to obtain agonist-bound β_1AR structures (*112*). Incorporation of the C-terminal hexahistidine tag allows purification of the receptor by Ni-NTA affinity chromatography.

Based on β_1AR44-m23 (*105*), three additional mutations I129V, D322K, and Y343L (Figure 3.1, blue residues) reported for the ultra-stable mutant β_1AR-JM3 (*110*) were added by site-directed mutagenesis. The point mutation I129V alone has been shown to increase the melting temperature of the receptor by about 2 °C. A more pronounced increase of about 5 °C has been achieved by the aspartate mutation D322K. Similarly, the point mutation Y343L has improved the thermal stability by 5 °C, which has been identified as the key stabilizing mutation for engineering of β_1AR-JM3 (*110*). Furthermore, the neutral mutation D200E was introduced, which is involved in the formation of a salt bridge to K322 and was reported to yield an increase in the melting temperature of β_1AR-JM3 in the short-chain detergent *n*-nonyl-β-D-glucopyranoside (NG) by 4.7 °C and by 6.2 °C in *n*-dodecyl-β-D-maltopyranoside (DDM) (*110*). Short-chain detergents are advantageous for solution state NMR due to a relatively small size of the solubilized receptor complex. However, β_1AR-JM3 solubilized in NG displayed a melting temperature of 35 °C, which was significantly lower compared to a melting temperature of 62 °C obtained for the DDM-solubilized receptor. Therefore, use of short-chain detergents is very likely not beneficial for NMR studies in solution.

When combined, the four mutations I129V, D200E, D322K, and Y343L were shown to increase the melting temperature of β_1AR44-m23 by 10 °C (*110*), which corresponds to an overall increase by 31 °C compared to the wild type sequence. However, these thermostabilizing mutations further reduce the binding affinity of TS-β_1AR for several agonists. In particular, K_i values determined for the agonist isoprenaline increased from about 17 nM in β_1AR44-m23 to 80 nM caused by the point mutations I129V, D322K and Y343L (*110*). This trend was similar for the agonist noradrenaline but less pronounced for the antagonists cyanopindolol and alprenolol (*110*).

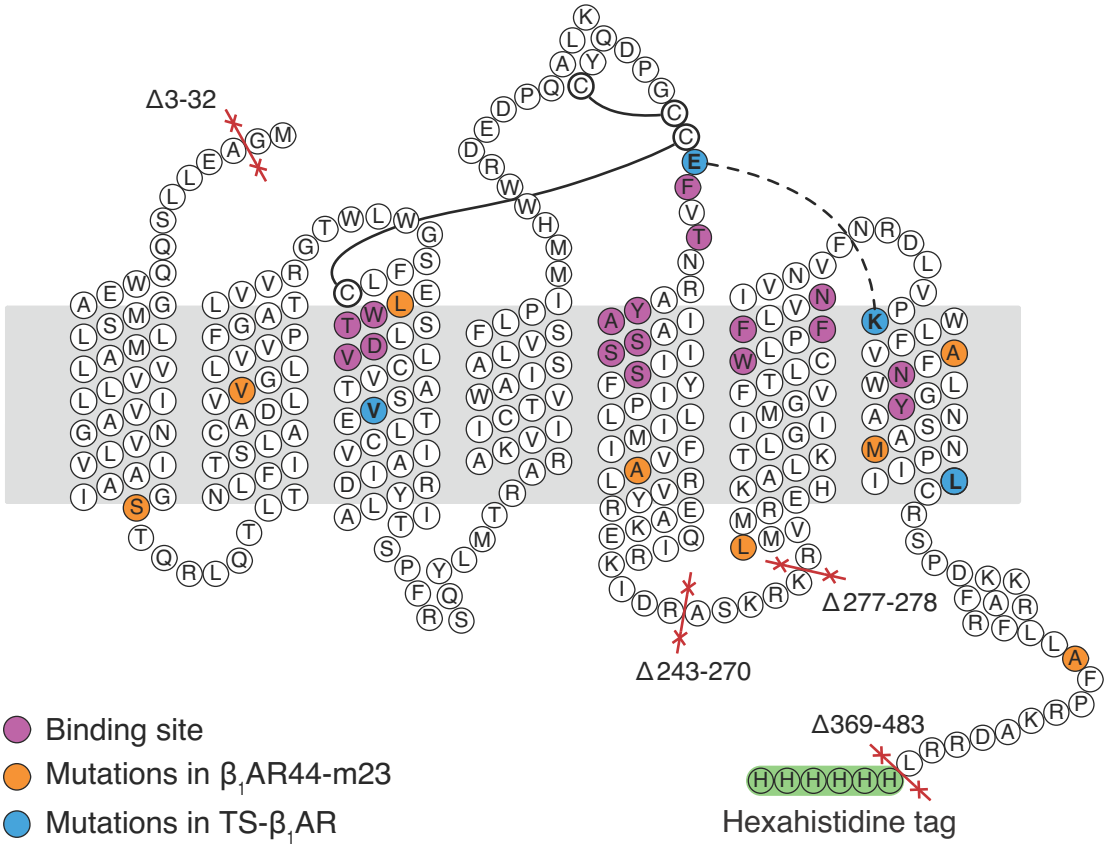


Figure 3.1 Topology and thermostabilization of β_1AR. Residues are color-coded as follows: binding site in purple, mutations in β_1AR44-m23 (*105*, *106*) in orange and mutations in TS-β_1AR in blue. Two distinct disulfide bridges are drawn as solid lines, whereas the salt bridge introduced by D322K and the neutral mutation D200E is shown as a dashed line. The red lines indicate truncations given by the residue numbers followed by the Δ.

Expression of TS-β_1AR in insect cells

The expression of TS-β_1AR in Sf9 insect cells was tested by preparation of the membrane fraction and subsequent solubilization in 1.5 % *n*-decyl-β-D-maltopyranoside (DM) without further purification. Figure 3.2 shows the separation of this solubilized fraction using a denaturing polyacrylamide gel electrophoresis (PAGE). However, identification of TS-β_1AR was hindered by the high background intensity, which was possibly caused by the high amount of DM. Western blot (WB) analysis revealed two prominent bands with an apparent molecular weight close to 25 kDa (Figure 3.2, WB) indicating "gel shifting" as compared to the expected 35.9 kDa, which is characteristic for membrane proteins and may originate from the interaction of the protein with sodium dodecyl sulfate (*113*). The reasons for the sample heterogeneity are currently unclear. The heterogeneity may result from degradation of the receptor by proteolysis. Typically, this is counteracted by using protease inhibitors during purification but may also be caused by the activity of proteases, which are insensitive to this treatment. Furthermore, as the posttranslational modifications sites of the wild type sequence have been removed in β_1AR44-m23 by the point mutation C358A of the single palmitoylation site (*105*, *106*) and truncation of the N-terminus (Δ3-32) to impede canonical *N*-linked glycosylation (*106*), sample heterogenity cannot be caused by these modifications.

However, the heterogeneity was eventually removed following purification by using an alprenolol-affinity chromatography step to select for active conformations of the receptor (*103*).

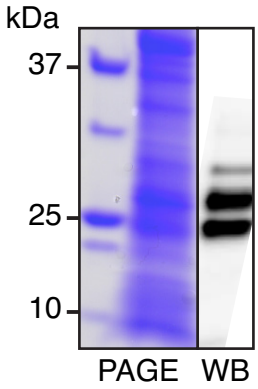


Figure 3.2 Expression of TS-β_1AR in Sf9 insect cells. Membrane fractions were solubilized in 1.5 % *n*-decyl-β-D-maltopyranoside and separated from insoluble material by centrifugation (see Materials and Methods). The soluble fraction was separated using SDS polyacrylamide gel electrophoresis (PAGE) and subsequently analyzed by Western blotting (WB). The first lane contains a molecular weight marker. TS-β_1AR was detected using an antibody specific for the C-terminal hexahistidine-tag.

Thermal shift assay of TS-β_1AR

To further characterize the engineered construct, thermal stability of TS-β_1AR expressed in insect cells was analyzed by a microscale fluorescent stability assay. This assay utilizes the thiol-specific fluorochrome N-[4-(7-diethylamino-4-methyl-3-coumarinyl)phenyl]maleimide (CPM) to probe temperature-induced unfolding of GPCRs in the presence of detergents (*114*). CPM contains a thiol-reactive maleimide moiety that covalently binds to free cysteine residues. Thermal unfolding exposes hydrophobic cysteine residues to the solvent, which are typically buried in the tightly packed inter-helix space of the receptor. Covalent binding to these exposed thiol groups forms intensely fluorescent thiol adducts, whereas the unbound molecule is essentially nonfluorescent (*115*).

Thermal unfolding of TS-β_1AR in 1 % *n*-decyl-β-D-maltopyranoside (DM) indicated a melting temperature of 57 °C (Figure 3.3), which is slightly higher than the melting temperature of 53 °C reported for β_1AR-JM3 (*110*). However, the determined melting temperatures vary considerably with respect to the utilized assay. Thermostability of β_1AR-JM3 has been analyzed using prolonged incubation times (30 min) at the desired temperatures (*110*). In contrast, the applied heating rate strongly influences the melting temperatures determined by the CPM assay. Furthermore, β_1AR-JM3 was shown to be 31 °C more thermostable compared to the wild type receptor (*110*). It is expected that a similar increase of the melting temperature can be assumed for TS-β_1AR based on the sequence homology of both stabilized receptors.

In addition, thermal stability was studied on a mutant of TS-β_1AR, for which the point mutation Y343L in the NPxxY motif of TM7 was reverted to the native tyrosine residue of the wild type receptor (TS-β_1AR-L343Y, Figure 3.3). The residue Y343 has been linked to G protein activation (*105*, *116*, *117*) and has been identified as a key residue for signal transduction in β_1AR only recently (*103*). Remarkably, the less stabilized mutant TS-β_1AR-L343Y solubilized in 1 % DM displayed a melting temperature of 53.5 °C, which is still considerably high for the apo receptor. Thermostabilization by Y343L may be related to the loss of a single water molecule close to the NPxxY motif as observed for the crystal structure of the ultra-stabilized β_1AR-JM3 (*106*, *110*).

Finally, the increase in the observed melting temperatures of TS-β_1AR-L343Y and TS-β_1AR relative to the wild type receptor agrees well with the trend of thermostabilization reported in recent crystallographic studies of β_1AR (*110*, *112*). The obtained improvement of the thermal stability is the basis for studies of these mutants by NMR.

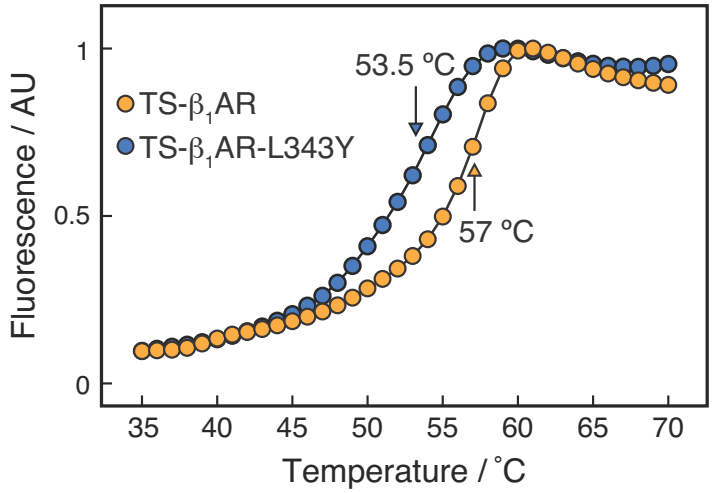


Figure 3.3 Temperature-induced unfolding of mutant receptors. Melting curves are shown for apo forms of TS-β_1AR and reverse mutant TS-β_1AR-L343Y obtained for 1 µg of receptor in 20 mM TRIS (pH 7.5), 0.1 % DM, and 350 mM NaCl. Thermal stability was probed by the thiol-specific fluorochrome CPM in a real-time PCR cycler using a heating rate of 2 K/min. The melting temperature was determined as the zero crossing of the computed first derivative of the melting curve (data provided by Shin Isogai).

Further development

In this section, a thermostabilized mutant (TS-β_1AR) of the β_1-adrenergic receptor has been derived based on previously utilized approaches for crystallization studies. Melting temperatures as high as 57 °C have been obtained.

The TS-β_1AR was used for an extensive NMR analysis described in the following section (Chapter 3.2). Labeling with ^{15}N-valine enabled backbone chemical shift analysis of the thermostabilized receptor to investigate the allosteric signaling network of the receptor and to test for G protein activation. The initial constructs of TS-β_1AR were then further used to assess the effect of thermostabilization on the biological activity of the receptor by reverting the introduced mutations towards the native sequence.

In the final section (Chapter 3.3), expression of ^{2}H,^{15}N-labeled TS-β_1AR in insect cells is described for backbone resonance assignments based on 3D ^{15}N-edited NOESY experiments. Very high spectral quality enabled the initial assignment of a consecutive stretch of residues, which is involved in signal transduction of the receptor upon ligand binding.

Material and Methods

Cloning of the thermostabilized β_1-adrenergic receptor

A pFastBac vector containing β_1AR44-m23 (*105*) was originally provided by Professor Gebhard Schertler (PSI, Villigen, Switzerland). Additional thermostabilizing mutations were

introduced by site-directed mutagenesis using the primers given in Table 3.1. The mutagenesis protocol was adapted from the commercial QuikChange method (Agilent).

Table 3.1 List of primers used for mutagenesis of the β_1-adrenergic receptor[a]

	Forward and reverse primer
I129V	5′-AAG C**GT** **G**GA GAC CTT GTG CGT CAT C 5′-GTC TC**C** **AC**G CTT GCC GTC ACG CAA A
D200E	5′-GCT GCG A**A**T TTG TCA CCA ACC GGG CTT 5′-ACA AA**T** TCG CAG CAG CCC GGG TCC T
D322K	5′-TGC CG**A** **AA**T GGC TCT TCG TTG CCT T 5′-AGC CA**T** **TT**C GGC ACC AGA TCT CTG TT
Y343L	5′-TCA TC**C** **T**CT GCC GCA GCC CAG ACT T 5′-CGG CAG **AG**G ATG ATG GGG TTC ATA G

[a] Bases modified with respect to the β_1AR44-m23 sequence are indicated in bold

The finale TS-β_1AR construct confirmed by Sanger DNA sequencing (MicroSynth AG, Switzerland) was

MGAELLSQQWEAGMSLLMAL VVLLIVAGNVLVIAAIGSTQ RLQTLTNLFITSLACADLVV
GLLVVPFGATLVVRGTWLWG SFLCELWTSLDVLCVTASVE TLCVIAIDRYLAITSPFRYQ
SLMTRARAKVIICTVWAISA LVSFLPIMMHWWRDEDPQAL KCYQDPGCCEFVTNRAYAIA
SSIISFYIPLLIMIFVALRV YREAKEQIRKIDRASKRKTS RVMLMREHKALKTLGIIMGV
FTLCWLPFFLVNIVNVFNRD LVPKWLFVAFNWLGYANSAM NPIILCRSPDFRKAFKRLLA
FPRKADRRLHHHHHH.

Cell culture and preparation of recombinant baculovirus

In general, handling and maintenance of Sf9 cell culture was based on established protocols (*118*). Permanent cultures were subcultured in mid-log phase and kept in full SF-4 medium (Bioconcept) and Insect Xpress medium (Lonza) at 27 °C and shaken at 80 rpm. Recombinant Bacmid DNA was generated in DH10Bac *E. coli* cells using the commercial pFastBac vector according to the supplier's procedure (Life Technologies). Baculovirus was generated in adherent cultures as reported by O'Reilly et al. (*118*). Subsequently, two additional amplification rounds in suspension cultures were employed to produce high-titer virus stocks for expression. In the first amplification round, 50 ml of Sf9 cell culture containing $1x10^6$ cells/ml in serum-free medium were infected with recombinant virus and cultured until cell viability decreased to ≤90 %. The supernatant (P1) was collected by centrifugation at 1000*g* and 4 °C for 15 min. For the second amplification round, 100 ml of serum-free Sf9 cell culture containing $1x10^6$ cells/ml was infected with P1 and processed accordingly to yield a P2 virus stock. The optimal P2 virus amount for expression was determined in small-scale cultures of Sf9 cells infected with P2 virus stock at a cell density of $1.5x10^6$ cells/ml in full SF-4 medium by varying the functional titer from 1 to 10 ml per 1 L of cell culture.

Expression and initial purification of TS-β_1AR

Expression was analyzed in membrane fractions prepared from small-scale cultures of Sf9 cells infected with recombinant baculovirus containing TS-β_1AR. Membranes were solubilized with 1.5 % *n*-decyl-β-D-maltopyranoside (Anatrace) in 20 mM TRIS (pH 7.8) and 400 mM NaCl and insoluble material was removed by ultracentrifugation at 50000*g* for 1 h (4 °C). The remaining soluble fraction was seperated by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). TS-β_1AR was detected by Western blotting using an antibody specific for the C-terminal hexahistidine tag. Complete purification of TS-β_1AR for NMR analysis is presented in Chapter 3.2.

Thermal shift assay of mutant receptors

Thermal stability analysis of TS-β_1AR and TS-β_1AR-L343Y was adapted from the recently established CPM-based microscale fluorescent assay (*114*). Briefly, melting curves were determined for 1 μg of purified apo receptor in a total volume of 65 μl in 20 mM TRIS (pH 7.5), 0.1 % DM, and 350 mM NaCl. Assays were performed in a real-time PCR cycler (Rotor-Gene Q, QIAGEN) using a ramp rate of 2 K/min. The melting temperature was determined as the inflection point of the melting curve by computing the first derivative.

3.2 Backbone NMR reveals allosteric signal transduction networks in the β_1-adrenergic receptor

Original publication

Shin Isogai, Xavier Deupi, Christian Opitz, Franziska M Heydenreich, Ching-Ju Tsai, Florian Brueckner, Gebhard F X Schertler, Dmitry B Veprintsev, and Stephan Grzesiek

Backbone NMR reveals allosteric signal transduction networks in the β_1-adrenergic receptor

Nature, 2016 vol. 530 (7589) pp. 237-241.

LETTER

doi:10.1038/nature16577

Backbone NMR reveals allosteric signal transduction networks in the β_1–adrenergic receptor

Shin Isogai[1], Xavier Deupi[2], Christian Opitz[1], Franziska M. Heydenreich[2], Ching-Ju Tsai[2], Florian Brueckner[2], Gebhard F. X. Schertler[2,3], Dmitry B. Veprintsev[2,3] & Stephan Grzesiek[1]

G protein-coupled receptors (GPCRs) are physiologically important transmembrane signalling proteins that trigger intracellular responses upon binding of extracellular ligands. Despite recent breakthroughs in GPCR crystallography[1–3], the details of ligand-induced signal transduction are not well understood owing to missing dynamical information. In principle, such information can be provided by NMR[4], but so far only limited data of functional relevance on few side-chain sites of eukaryotic GPCRs have been obtained[5–9]. Here we show that receptor motions can be followed at virtually any backbone site in a thermostabilized mutant of the turkey β_1-adrenergic receptor (β_1AR)[10–12]. Labelling with [^{15}N]valine in a eukaryotic expression system provides over twenty resolved resonances that report on structure and dynamics in six ligand complexes and the apo form. The response to the various ligands is heterogeneous in the vicinity of the binding pocket, but gets transformed into a homogeneous readout at the intracellular side of helix 5 (TM5), which correlates linearly with ligand efficacy for the G protein pathway. The effect of several pertinent, thermostabilizing point mutations was assessed by reverting them to the native sequence. Whereas the response to ligands remains largely unchanged, binding of the G protein mimetic nanobody NB80 and G protein activation are only observed when two conserved tyrosines (Y227 and Y343) are restored. Binding of NB80 leads to very strong spectral changes throughout the receptor, including the extracellular ligand entrance pocket. This indicates that even the fully thermostabilized receptor undergoes activating motions in TM5, but that the fully active state is only reached in presence of Y227 and Y343 by stabilization with a G protein-like partner. The combined analysis of chemical shift changes from the point mutations and ligand responses identifies crucial connections in the allosteric activation pathway, and presents a general experimental method to delineate signal transmission networks at high resolution in GPCRs.

A thermostabilized, detergent-resistant mutant of β_1AR[10–12] (TS-β_1AR; see Methods) was selectively labelled with [^{15}N]valine and produced in insect cells without further chemical modifications. Its 28 valine residues are homogeneously distributed across the receptor (Extended Data Fig. 1) at locations suitable to sense ligand binding and receptor activation. Although resonances of main chain atoms are considerably more difficult to observe than those of mobile side chains of surface residues, they are expected to be better reporters of functional, long-range backbone motions. We succeeded to obtain well-resolved TROSY (transverse relaxation-optimized spectroscopy) spectra of the valine ^{1}H–^{15}N backbone resonances of detergent-solubilized TS-β_1AR in its apo form and in complexes with six ligands (Extended Data Fig. 1) ranging in their efficacy from antagonists to agonists (Extended Data Table 1). Despite the absence of deuteration and very short T_2 relaxation times (~4 ms for $^1H^N$), 26 valine resonances could be observed with sufficient sensitivity and resolution. Distinct and reversible chemical shift changes were detected for many valines after ligand exchange. 16 valines were assigned unambiguously and 5 tentatively using spectra from 18 point mutants, as well as further spectral and structural information (Extended Data Table 2).

Many valine residues in the vicinity of the ligand binding pocket could be assigned, showing chemical shift changes that report on the ligand functional groups (Fig. 1). Remarkably, residue V172(4.56) (the number in parenthesis corresponds to the Ballesteros–Weinstein numbering system[13]), which is located close to the ligand aromatic head group, exhibits an unusual ^{15}N chemical shift of ~105–110 p.p.m. (Extended Data Figs 1, 2 and Fig. 1c). This anomaly seems caused by a distorted backbone geometry, which is presumably conserved among adrenergic receptors and results from a missing hydrogen bond to the proline at position 176(4.60) (Extended Data Fig. 2). Instead, the carbonyl of V172(4.56) participates in a water-mediated hydrogen bond network, which connects the ligand binding site, TM3, TM4, TM5, and TM6[12,14]. Seemingly as a result of these interactions, the V172(4.56) ^{1}H–^{15}N resonances cluster according to the substitution patterns of the ligand head group (Fig. 1c): one cluster is observed for the partial agonists/antagonists cyanopindolol, alprenolol and carvedilol, which have larger head groups with *ortho*- and/or *meta*-substitutions; a second cluster is observed for the agonists isoprenaline and dobutamine, which bear a *meta*- and *para*-substituted catechol ring. We attribute the distinct chemical shifts for isoprenaline or dobutamine to the loss of a coordinated water caused by specific hydrogen bond interactions between their catechol moieties and the side chain of S215(5.461) (Extended Data Fig. 2d).

Remarkably, the resonance positions of V172(4.56) in complex with the antagonist atenolol strongly differ from the already described complexes and the apo form: considerable $^1H^N$ (>0.4 p.p.m.) and ^{15}N (>4 p.p.m.) upfield shifts indicate, respectively, a further weakening of the main chain hydrogen bond V172(4.56)H^N···I168(4.52)O and a stronger kink of the backbone. This rearrangement seems caused by the insertion of the *para*-acetamide group of the ligand head between residues S215(5.461) and V172(4.56) (Extended Data Fig. 2d). This will lead to a substantial disruption of the TM3–TM4–TM5 interface, thereby precluding receptor activation, in agreement with atenolol's inverse agonist pharmacology. Thus, the amide chemical shifts of V172(4.56) constitute a very sensitive readout for the state of this water-mediated, inter-helical activation switch.

The ^{1}H–^{15}N chemical shifts of further residues in the vicinity of the binding pocket report on additional characteristics of the ligands. V314(6.59) and V202(ECL2) are located at the extracellular surface of the receptor in a "vestibule" next to the entry/exit pathway of the orthosteric binding site[15,16]. The resonances of these residues are either severely broadened or undetectable in the absence of ligands, whereas they are observable in the presence of ligands (Fig. 1d and Extended Data Fig. 3). A line shape analysis for V314(6.59) (Extended Data Fig. 4)

[1]Focal Area Structural Biology and Biophysics, Biozentrum, University of Basel, CH-4056 Basel, Switzerland. [2]Paul Scherrer Institute, CH-5232 Villigen PSI, Switzerland. [3]Department of Biology, ETH Zurich, CH-8093 Zurich, Switzerland.

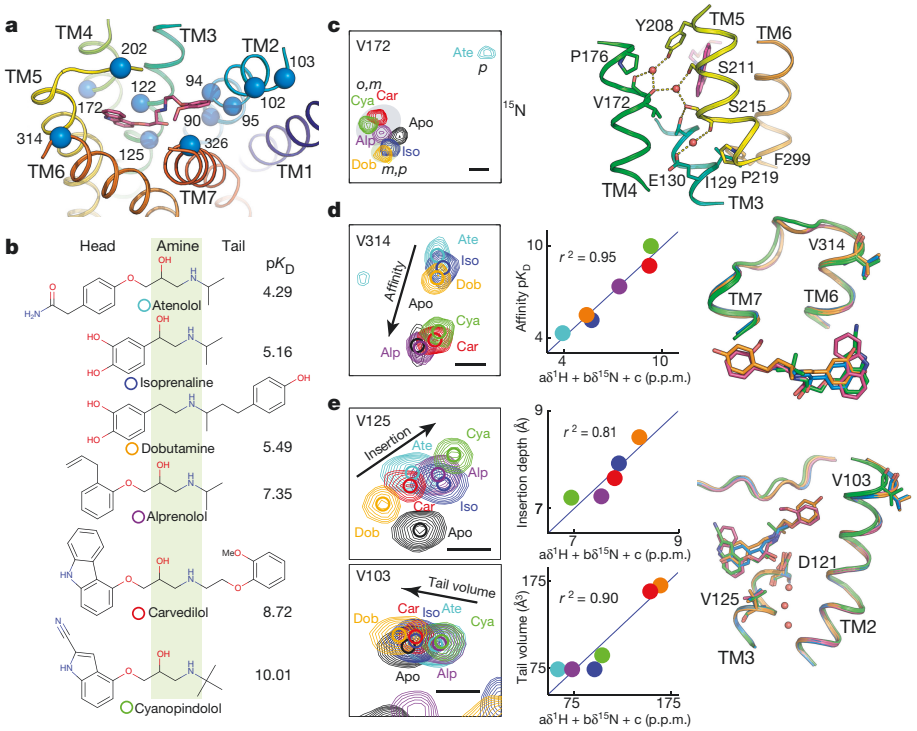


Figure 1 | Ligand-induced ^{1}H–^{15}N chemical shift changes in the vicinity of the ligand binding pocket of β_1AR. a, Partial view of the β_1AR–carvedilol crystal structure (4AMJ) showing valine residues (blue spheres) in the vicinity (<8.5 Å) of the ligand (magenta sticks) binding site. **b**, Chemical structures of the β_1AR ligands used in this study. Ligand affinities derived from whole-cell binding assays on the thermostabilized β36-m23 β_1AR construct[17] are indicated as pK_D values. Similar pK values were measured for the TS-β_1AR construct (Extended Data Table 1). **c**, Left, ligand-induced response of V172(4.56) ^{1}H–^{15}N resonances. The black bar represents a scale of 0.1 p.p.m. (^{1}H) and 1 p.p.m. (^{15}N). The labels *o*, *m*, *p* indicate ligands with *ortho*, *meta*, and *para* substitutions at the head group, respectively. **c**, Right, partial view of the β_1AR–carvedilol structure (4AMJ) showing the interaction network connecting V172(4.56) to S215(5.461), P219(5.50), I129(3.40), and F299(6.44). **d**, Left, representation as **c**, left, for the ^{1}H–^{15}N resonances of V314(6.59). Centres of resonances are indicated by circles. **d**, Middle, correlation of a best-fit linear combination of the V314(6.59) chemical shifts ($-49.3\ \delta^{1}H + 2.02\ \delta^{15}N + 133$) to the ligand affinity pK_D (Extended Data Table 1). **d**, Right, partial view of the carvedilol (red, 4AMJ), dobutamine (orange, 2Y01), cyanopindolol (green, 2VT4 B), and isoprenaline (blue, 2Y03) complex structures showing the ligand-induced movement of V314(6.59). **e**, Left, representation as **d**, left, for the ^{1}H–^{15}N resonances of V125(3.36) and V103(2.65). **e**, Middle, correlations of best-fit linear combinations of chemical shifts for V125(3.36) ($-0.402\ \delta^{1}H + 1.17\ \delta^{15}N - 132$) and V103(2.65) ($-336\ \delta^{1}H - 634\ \delta^{15}N + 7.64{\cdot}10^{4}$) to the ligand insertion depths and tail volumes (Extended Data Table 1), respectively. **e**, Right, partial view of the carvedilol, dobutamine, cyanopindolol, and isoprenaline crystal structures (representation as **d**, right) showing the ligand-induced movement of V125(3.36), D121(3.32), and V103(2.65).

indicates that this extracellular part of the receptor undergoes micro- to millisecond motions in the apo form, which are quenched by ligand binding. This is consistent with results on β_2AR, which suggest that high-affinity ligands stabilize the conformation of ECL2 and ECL3[5]. The ligand-induced shifts of the V314 resonance correlate strongly ($r^2 = 0.95$) with the reported ligand affinity[17] (Fig. 1d). Interestingly, the resonances of V314(6.59) in the apo form and in high-affinity ligand complexes are very close. This may indicate that the high-affinity ligand complexes mimic the average apo conformation. Finally, the ^{1}H–^{15}N chemical shifts of V125(3.36) at the bottom of the binding site and of V103(2.65) close to the ligand tail reveal additional trends (Fig. 1e): the chemical shifts of V125(3.36) correlate with the depth of ligand insertion towards the central part of TM3 ($r^2 = 0.81$), whereas those of V103 correlate with the volume of the ligand tail ($r^2 = 0.90$).

Compared to inactive β_2AR, complexes of activated β_2AR with either G protein[2] or the G protein-mimicking nanobody NB80[9,18] show large movements at the intracellular sides of TM5, TM6 and their intervening loop ICL3, which form the binding site for the G protein. These conformational changes are expected to be conserved throughout the GPCR family[19]. Four valine residues could be assigned in this region of TS-β_1AR: V226(5.57), V230(5.61), V280(6.25), and V298(6.43) (Fig. 2). In contrast to the chemical shift changes in the vicinity of the ligand binding pocket, which depend strongly on the ligand chemistry, the shifts of the TM5 residues observed in this region report on ligand efficacy. This effect is most prominent for residue V226(5.57), for which the ^{1}H–^{15}N resonances fall on one line from antagonists to agonists (Fig. 2a). The chemical shifts for the different ligands correlate very strongly ($r^2 = 0.89$) with their reported[17] efficacies for Gs signalling (Extended Data Table 1). This highly linear effect suggests that the receptor filters the diverse input signals from the various ligands to a unified and precise structural response on TM5, which can be read out by the chemical shifts of V226(5.57). Interestingly, the V226(5.57) atenolol peak is situated at a position corresponding to lower efficacy than for the apo receptor. This gives direct structural evidence of atenolol's inverse agonist action, which reduces the activation relative to the basal level of the apo receptor.

Current high-resolution structures of β_1AR do not show significant changes between antagonist- and agonist-bound forms (Fig. 2b). The decrease of the V226(5.57) $^{1}H^{N}$ chemical shift by about 0.2 p.p.m. from the agonist isoprenaline to the antagonist atenolol indicates a lengthening of the V226(5.57)-H^{N}···I222(5.53)-O hydrogen bond by about 0.05 Å (ref. 20). This small, but clearly NMR-detectable length variation is below the resolution limit of current GPCR structures (Supplementary Text 1), but may indicate the start of TM5 bending towards the active conformation as observed in the G protein-bound form of β_2AR (Fig. 2b). Remarkably, this response to agonists occurs even in the thermostabilized receptor TS-β_1AR. Albeit reduced in absolute size compared to V226(5.57), V230(5.61) displays similar linear chemical shift changes as a function of ligand efficacy for the

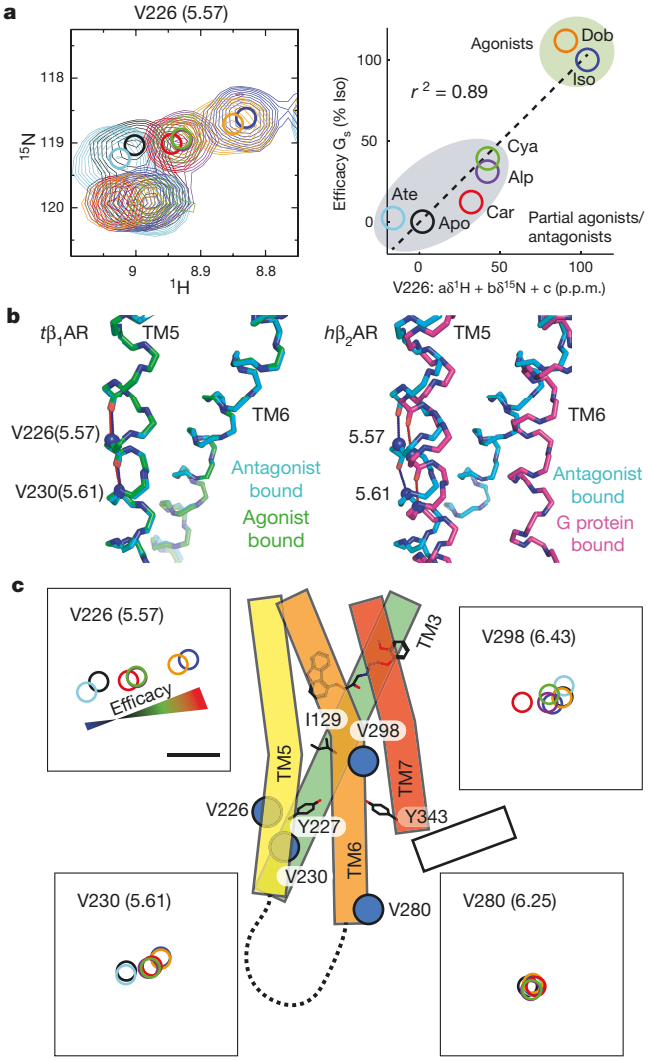


Figure 2 | Correlation of ligand-induced chemical shift changes at the TS-β_1AR intracellular side with G_s efficacy. a, Left, response of the V226(5.57) ^{1}H–^{15}N resonance to various ligands (colour coding as in Fig. 1). The centres of resonances are indicated by circles. The ^{1}H–^{15}N resonances fall on one line from atenolol (antagonist) over apo to alprenolol (partial agonist), carvedilol (antagonist), cyanopindolol (partial agonist), dobutamine (full agonist) and isoprenaline (full agonist). **a**, Right, correlation of a best-fit linear combination of the V226(5.57) chemical shifts (-515 δ^1H -31.7 δ^{15}N $+8.41{\cdot}10^3$) in different ligand complexes to their efficacy for the Gs signalling pathway[17]. **b**, Left, overlay of TM5 and TM6 backbones of thermostabilized β_1AR in antagonist- (blue, PDB code 4AMJ) and agonist-bound (green, PDB code 2Y03) form. The agonist does not induce detectable helix movements. **b**, Right, TM5 and TM6 backbone movements upon activation in human β_2AR. The overlay of inactive (blue, PDB code 2RH1) and G protein-bound β_2AR (magenta, PDB code 3SN6) structures shows the large bend of TM6 along with the smaller conformational change of TM5 upon activation. Hydrogen bonds 5.57-H^N···5.53-O and 5.61-H^N···5.57-O are indicated by dashes. According to the behaviour of the ^{1}H^N chemical shifts of V226(5.57) and V230(5.61) in TS-β_1AR, these hydrogen bonds expand in an efficacy-dependent manner during agonist binding. **c**, Response of the ^{1}H–^{15}N resonances for V226(5.57), V230(5.61), V298(6.43), and V280(6.25) to various ligands. For clarity, only the centres of resonances are depicted as circles with colour coding as in Fig. 1. The black bar represents a scale of 0.1 p.p.m. and 1 p.p.m. for the ^{1}H and ^{15}N chemical shifts, respectively. The schematic representation of the receptor indicates the locations of the respective valine residues at the cytoplasmic sides of TM5 and TM6 within the helical bundle of β_1AR.

G_s pathway (Fig. 2c). As this residue is located one helical turn further towards the cytoplasm, the detected conformational change is not just local, but spans a certain length in TM5.

Compared to V226(5.57) and V230(5.61) in TM5, the chemical shift response to ligands is much less pronounced for V298(6.43) and in particular V280(6.25) at the intracellular side of TM6 (Fig. 2c). This suggests that agonist binding to the TS-β_1AR does not induce the large conformational change in TM6 observed in the activated β_2AR–G protein[2] or β_2AR–NB80[18] complexes. However, G protein activation upon agonist binding has been reported for other less thermostabilized β_1AR constructs[17], indicating that they can still be activated, albeit at low levels. Thus we reverted the mutations most likely to interfere with the activation mechanism in TS-β_1AR, that is, I129(3.40)V in the connector switch[15], Y227(5.58)A in TM5[21,22], and Y343(7.53)L in the NPxxY motif of TM7[14,22] to the native residues. These reverse mutants were then tested for G protein activation and the NMR response in TM6. A summary of the results is given in Extended Data Table 3. The original TS-β_1AR, the single mutants TS-β_1AR(V129I), TS-β_1AR(A227Y), and TS-β_1AR(L343Y) as well as the double mutant TS-β_1AR(V129I/A227Y) showed no detectable G protein activation upon isoprenaline binding (Extended Data Fig. 5). However, G protein activation was detectable for the least thermostable TS-β_1AR(A227Y/L343Y) double mutant (T_m reduced by 11 °C relative to TS-β_1AR), which recovers the conserved tyrosines in TM5 and TM7 that are known to stabilize the active state of rhodopsin[22]. None of the reverse mutants showed major changes in the NMR spectra of various ligand complexes compared to the original TS-β_1AR (Extended Data Fig. 6). In particular, residues V298(6.43) and V280(6.25) at the intracellular side of TM6 did not show an increased response to agonists. This is in agreement with recent DEER (double electron–electron resonance) and ^{19}F-NMR data showing that agonists alone do not fully stabilize the active state of TM6 in β_2AR[9]. Interestingly, the V129(3.40)I and the V129(3.40)I/A227(5.58)Y mutations shifted the ^{1}H–^{15}N resonances of V226(5.57) towards a more active (that is, bent) state of TM5 in both the atenolol- and isoprenaline-bound forms (Extended Data Fig. 6b), thereby given direct experimental evidence for an allosteric activation pathway spanning about 13 Å from I129(3.40) on TM3 to V226(5.57) on TM5.

With the exception of rhodopsin[23], the stabilization of fully active GPCR conformations seems to require binding of an agonist and an intracellular partner[8,9]. Indeed, when both the agonist isoprenaline and the G protein-mimicking nanobody NB80[18] were added to TS-β_1AR(A227Y/L343Y), very large chemical shift responses for many valine residues in TM3–TM6 were observed, whereas no change was observed for several valines in TM1, 2, and 7 (Extended Data Fig. 7). This very strong response extends even to the extracellular residue V314(6.59), providing evidence of a long-distance connection from the G protein binding site to the ligand entry site. The strong chemical shift changes are reverted when the partial agonist/antagonist cyanopindolol is added to the isoprenaline-TS-β_1AR(A227Y/L343Y)–NB80 complex. The spectrum then becomes identical to that of the 'pure' cyanopindolol–TS-β_1AR(A227Y/L343Y) complex (Extended Data Fig. 7a), indicating that cyanopindolol replaces isoprenaline and causes the release of NB80. In agreement with the G protein activation data, the isoprenaline-bound original TS-β_1AR and the mutants TS-β_1AR(A227Y) and TS-β_1AR(L343Y) did not show binding of NB80 in the NMR spectra. Moreover, supplementing NB80 to the ultrastable TS-β_1AR did not change its affinity for isoprenaline (Extended Data Fig. 7b), whereas it caused a hundred-fold affinity increase in the case of TS-β_1AR(A227Y/L343Y) and the truncated native turkey β_1AR receptor (tβtrunc)[17]. This increase is identical to data for β_2AR[18] and shows the energetic coupling between the NB80 and agonist binding also for β_1AR.

In combination, these data prove that agonist binding, even in the absence of a G protein mimic, induces initial changes in the conformational equilibrium of TM5 towards the conformation observed in the G protein complex of β_2AR. Remarkably, these rearrangements occur in all thermostabilized forms of β_1AR. However, a full shift of the equilibrium towards such an active conformation, including allosteric

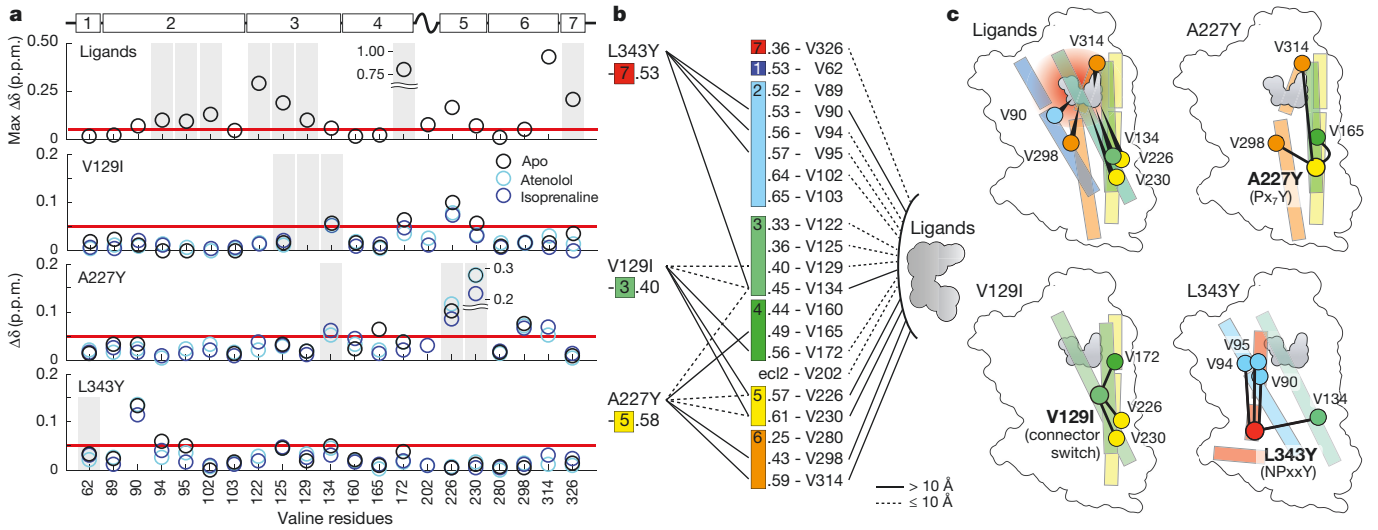


Figure 3 | Experimental detection of allosteric signalling pathways using the NMR response to ligand binding and point mutations at different backbone sites. a, Combined ^{1}H, ^{15}N chemical shift deviations $[\Delta\delta = (\Delta\delta_{1H}^2/2 + \Delta\delta_{15N}^2/50)^{1/2}]$ of valine resonances observed upon ligand binding or induced by the indicated point mutations. For ligand binding, the three pairwise deviations $\Delta\delta$ were calculated between the apo, atenolol-, and isoprenaline-bound forms of TS-β_1AR. The maximum of these deviations is shown. For the reverse mutants, deviations $\Delta\delta$ are shown relative to TS-β_1AR for their apo (black), atenolol- (cyan), and isoprenaline-bound (blue) bound forms. Valines within 10 Å from the Cα atom of the mutated amino acid are shown on a grey background. Distances were calculated using the coordinates of the thermostabilized β_1AR (PDB code 4BVN). A red line marks a cut-off value $\Delta\delta$ of 0.05 p.p.m. for significant chemical shift deviations. **b**, Topology of the signalling network determined from point mutations (left) and ligand binding (right). Signal paths were identified by chemical shift deviations $\Delta\delta$ larger than 0.05 p.p.m. induced by these two perturbations (**a**). Signal paths to valines within 10 Å from the ligand or point mutation (that is, localized conformational changes) are indicated as dashed lines, and those beyond 10 Å (long-range conformational changes) as solid lines. The ligand signals broadly towards all helices but TM1. In contrast, the network determined by the point mutations is more localized and connects TM3 to TM4/5, TM5 to TM3/4/6, and TM7 to TM2/3. The latter network seems to be divided into two subnetworks involving TM3/4/5/6 and TM2/7. **c**, Long-range allosteric signal paths identified from ligand binding or point mutations (**a** and **b**) indicated on schematic β_1AR representations showing the involved TMs. Helices are colour-coded according to **b**.

changes at the extracellular side, occurs only when G protein or its mimetic NB80 is bound. This process requires the presence of both Y227(5.58) and Y343(7.53), which significantly reduce the thermal stability. Different active conformations may be reached for non-G protein effectors such as β-arrestin.

The possibility to detect NMR signals at many receptor sites in response to ligand binding and point mutations provides an experimental method to trace allosteric signalling paths. Figure 3 shows examples of these pathways, derived from the response to the ligands atenolol and isoprenaline and the single point mutations V129(3.40)I, A227(5.58)Y, and L343(7.53)Y. Choosing a cutoff of 0.05 p.p.m. for the resulting combined ^{1}H, ^{15}N chemical shift change (Fig. 3a, red line), long-range (>10 Å) connections become evident throughout the receptor (Fig. 3b, c). Whereas detected ligand signals radiate broadly to almost all helices, the point mutants give evidence of smaller interaction networks connecting TM3 to TM4/5, TM5 to TM3/4/6, as well as TM7 to TM2/3 (Fig. 3b, c). Interestingly, the TM2/TM7 network seems to be only weakly connected to the TM3–TM6 network. Together, these data provide experimental evidence at high resolution of an extensive signal transduction network that connects the ligand binding site to the intracellular sides of TM5, TM6, and TM7. Such a network of loosely coupled allosteric connections has been postulated previously for β_2AR on the basis of molecular dynamics simulations[24].

In summary, we have shown that highly resolved solution NMR backbone spectra can be obtained for a eukaryotic GPCR. The NMR observations delineate the allosteric signalling pathways and comprehensively connect many previous experimental and theoretical observations, which may ultimately allow to understand the dynamic mechanisms of GPCRs at the atomic level.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 16 October 2014; accepted 4 December 2015.
Published online 3 February 2016.

1. Rasmussen, S. G. F. *et al.* Crystal structure of the human β_2 adrenergic G-protein-coupled receptor. *Nature* **450,** 383–387 (2007).
2. Rasmussen, S. G. F. *et al.* Crystal structure of the β_2 adrenergic receptor-Gs protein complex. *Nature* **477,** 549–555 (2011).
3. Ghosh, E., Kumari, P., Jaiman, D. & Shukla, A. K. Methodological advances: the unsung heroes of the GPCR structural revolution. *Nature Rev. Mol. Cell Biol.* **16,** 69–81 (2015).
4. Grzesiek, S. & Sass, H.-J. From biomolecular structure to functional understanding: new NMR developments narrow the gap. *Curr. Opin. Struct. Biol.* **19,** 585–595 (2009).
5. Bokoch, M. P. *et al.* Ligand-specific regulation of the extracellular surface of a G-protein-coupled receptor. *Nature* **463,** 108–112 (2010).
6. Kofuku, Y. *et al.* Efficacy of the β_2-adrenergic receptor is determined by conformational equilibrium in the transmembrane region. *Nat. Commun.* **3,** 1045–1049 (2012).
7. Liu, J. J., Horst, R., Katritch, V., Stevens, R. C. & Wüthrich, K. Biased signaling pathways in β_2-adrenergic receptor characterized by ^{19}F-NMR. *Science* **335,** 1106–1110 (2012).
8. Nygaard, R. *et al.* The dynamic process of β_2-adrenergic receptor activation. *Cell* **152,** 532–542 (2013).
9. Manglik, A. *et al.* Structural insights into the dynamic process of β_2-adrenergic receptor signaling. *Cell* **161,** 1101–1111 (2015).
10. Warne, T. *et al.* Structure of a β_1-adrenergic G-protein-coupled receptor. *Nature* **454,** 486–491 (2008).
11. Miller, J. L. & Tate, C. G. Engineering an ultra-thermostable β_1-adrenoceptor. *J. Mol. Biol.* **413,** 628–638 (2011).
12. Miller-Gallacher, J. L. *et al.* The 2.1 Å resolution structure of cyanopindolol-bound β1-adrenoceptor identifies an intramembrane Na^+ ion that stabilises the ligand-free receptor. *PLoS One* **9,** e92727 (2014).
13. Ballesteros, J. A. & Weinstein, H. Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. *Methods Neurosci.* **25,** 366–428 (1995).
14. Deupi, X. & Standfuss, J. Structural insights into agonist-induced activation of G-protein-coupled receptors. *Curr. Opin. Struct. Biol.* **21,** 541–551 (2011).
15. Dror, R. O. *et al.* Pathway and mechanism of drug binding to G-protein-coupled receptors. *Proc. Natl Acad. Sci. USA* **108,** 13118–13123 (2011).

16. González, A., Perez-Acle, T., Pardo, L. & Deupi, X. Molecular basis of ligand dissociation in β-adrenergic receptors. *PLoS One* **6,** e23815 (2011).
17. Baker, J. G., Proudman, R. G. W. & Tate, C. G. The pharmacological effects of the thermostabilising (m23) mutations and intra and extracellular (β36) deletions essential for crystallisation of the turkey β-adrenoceptor. *Naunyn Schmiedebergs Arch. Pharmacol.* **384,** 71–91 (2011).
18. Rasmussen, S. G. F. *et al.* Structure of a nanobody-stabilized active state of the β_2 adrenoceptor. *Nature* **469,** 175–180 (2011).
19. Manglik, A. & Kobilka, B. The role of protein dynamics in GPCR function: insights from the β_2AR and rhodopsin. *Curr. Opin. Cell Biol.* **27,** 136–143 (2014).
20. Grzesiek, S., Cordier, F., Jaravine, V. & Barfield, M. Insights into biomolecular hydrogen bonds from hydrogen bond scalar couplings. *Prog Nucl Mag Res Sp* **45,** 275–300 (2004).
21. Tate, C. G. & Schertler, G. F. Engineering G protein-coupled receptors to facilitate their structure determination. *Curr. Opin. Struct. Biol.* **19,** 386–395 (2009).
22. Goncalves, J. A. *et al.* Highly conserved tyrosine stabilizes the active state of rhodopsin. *Proc. Natl Acad. Sci. USA* **107,** 19861–19866 (2010).
23. Park, J. H., Scheerer, P., Hofmann, K. P., Choe, H.-W. & Ernst, O. P. Crystal structure of the ligand-free G-protein-coupled receptor opsin. *Nature* **454,** 183–187 (2008).
24. Dror, R. O. *et al.* Activation mechanism of the β_2-adrenergic receptor. *Proc. Natl Acad. Sci. USA* **108,** 18684–18689 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We acknowledge T. Sharpe for expert help with biophysical assays. We thank J. Steyaert for providing the NB80 plasmid and M. Rogowski for preparing wild-type β_1AR baculovirus. F.B. was supported by Marie Curie and EMBO postdoctoral fellowships. This work was supported by Swiss National Science Foundation grants 31-132857 (S.G.), Sinergia 141898 (S.G., D.B.V., G.F.X.S.), 31-135754 (D.B.V.), 31-153145 (G.F.X.S.), 31-146520 (X.D.), European Union FP7 grant 242135 (S.G.), and COST Action CM1207 (GLISTEN) (X.D. and G.F.X.S.).

Author Contributions S.G., G.F.X.S., D.B.V., X.D. and S.I. initiated and managed the project. S.I., X.D., F.B., D.B.V., G.F.X.S. and S.G. designed ligand response and selective labelling experiments. C.O. designed the initial TS-β_1AR construct and established purification. S.I. designed and prepared all selectively labelled receptor mutants, performed ligand exchange, NB80 binding, and all NMR experiments. S.I. and S.G. designed NMR experiments, analysed and interpreted all data. F.M.H. and D.B.V. purified trimeric G_i protein, designed and performed radioligand affinity and G_i protein activation assays. C.-J.T. prepared NB80 nanobody. S.G., S.I., X.D., F.M.H., D.B.V. and G.F.X.S. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.G. (stephan.grzesiek@unibas.ch), D.B.V. (dmitry.veprintsev@psi.ch) or G.F.X.S. (gebhard.schertler@psi.ch).

METHODS

β_1AR constructs. The TS-β_1AR mutant was derived from the turkey β_1AR44-m23 mutant used in crystallographic studies[25] by adding three additional thermostabilizing mutations (I129V, D322K, and Y343L) and a neutral mutation (D200E) from the ultra-stable β_1AR-JM3 mutant[11]. As compared to the wild type, TS-β_1AR contains truncations at the amino and carboxy termini and intracellular loop (ICL3), a total of nine thermostabilizing point mutations, three further point mutations as well as a C-terminal hexahistidine tag (Extended Data Fig. 1). The final TS-β_1AR sequence is MGAELLSQQWEAGMSLLMALVVLLIVAGNVLVIAAIGSTQRLQTLTN LFITSLACADLVVGLLVVPFGATLVVRGTWLWGSFLCELWTSLDVLCVTASVE TLCVIAIDRYLAITSPFRYQSLMTRARAKVIICTVWAISALVSFLPIMMHWWR DEDPQALKCYQDPGCCEFVTNRAYAIASSIISFYIPLLIMIFVALRVYREAKEQ IRKIDRASKRKTSRVMLMREHKALKTLGIIMGVFTLCWLPFFLVNIVNVFNR DLVPKWLFVAFNWLGYANSAMNPIILCRSPDFRKAFKRLLAFPRKADRRLH HHHHH.

Additional valine-to-alanine or isoleucine point mutations were introduced into TS-β_1AR for NMR assignment purposes. All constructs were made using the QuikChange site-directed mutagenesis method (Agilent). Baculovirus for insect cell expression was generated using the Bac-to-Bac system (Invitrogen).

β_1AR expression and purification. All β_1AR constructs were expressed in baculovirus-infected insect cells as described[26]. Selective labelling by [^{15}N]valine was achieved by growing cells on unlabelled serum-free insect cell medium (InsectXpress, Lonza) and then exchanging into custom-made serum-free medium (SF4, BioConcept) devoid of valine and yeast extract, to which 100 mg l^{-1} [^{15}N] valine were supplemented. Virus was added immediately after the medium exchange. The culture was harvested at 48 or 72 h post infection.

After cell lysis, the membrane fraction was separated from the lysate via ultracentrifugation and subsequently solubilized with 2% *n*-decyl-β-D-maltopyranoside (DM, Anatrace). The solubilized membrane fraction was then purified by nickel ion affinity chromatography followed by alprenolol ligand affinity chromatography. The active receptor was eluted with buffer (20 mM Tris, 350 mM NaCl, 0.1% DM, pH 7.5) containing either atenolol (1 mM) or alprenolol (0.1 mM). Final yields of detergent-solubilized receptor were 1.5 mg l^{-1} of cell culture. The molecular weight of the receptor-detergent complex was estimated as ~100 kDa by static light scattering.

Thermal shift assays of mutant receptors. Detergent-solubilized, purified apo TS-β_1AR and reverse-mutation receptors for thermal stability assays were obtained from their atenolol-bound form by washing with buffer devoid of ligand on a HiTrap SP HP (GE Healthcare) column. Their thermal stability was determined by the microscale fluorescent stability assay for binding of the thiol-specific fluorochrome N-[4-(7-diethylamino-4-methyl-3-coumarinyl)phenyl] maleimide (CPM)[27] in a Rotor-Gene Q (QIAGEN) real-time PCR cycler using 1 μg of receptor in 20 mM Tris, 350 mM NaCl, 0.1% DM, pH 7.5 and a heating rate of 2 K min^{-1}.

NMR experiments. NMR samples were prepared in Shigemi tubes as 250 μl volumes of typically 100 μM receptor, 1 mM ligand (except for apo form), 20 mM TRIS, 100 mM NaCl, 0.1% DM, 5% D_2O, pH 7.5. For isoprenaline or dobutamine, 2 mM of sodium L-ascorbate were supplemented as anti-oxidant. All solution NMR measurements were carried out at 304 K on a 800 MHz or a 900 MHz Bruker Avance III spectrometer equipped with a cryogenic probe. 2D ^{1}H,^{15}N TROSY spectra were recorded with total acquisition periods of 16 ms (^{15}N) and 43 ms (^{1}H) with typical total experimental times of 24–48 h. As compared to a standard TROSY pulse sequence, the ^{1}H–^{15}N INEPT delays were set to 3 ms to reduce magnetization losses from relaxation.

Assignment procedure. To obtain sequence-specific assignment information, we initially attempted to detect HNCO and HNCA correlations on samples additionally labelled with ^{13}C at specific backbone sites[28]. However, due to low sensitivity, only very few correlations were observable. Therefore, assignments were obtained from a combination of information from TROSY spectra recorded on 18 TS-β_1AR valine point mutants with different ligands, four HN(CO) correlations and five distinct structure-based chemical shift predictions (Extended Data Table 2).

Ligand exchange experiments. Receptor complexes with different ligands were generated by sequential exchange according to increasing ligand affinity, that is, in the sequence atenolol-isoprenaline-dobutamine-alprenolol or alprenolol-carvedilol-cyanopindolol. For exchange, the sample was washed three times with buffer devoid of ligand at tenfold dilution in Amicon Ultra 50 kDa cutoff concentrators. Subsequently, the sample was washed again twice with buffer containing 100 μM new ligand, separated by a period of 1 h incubation. Final concentrations of the ligands were adjusted to 1 mM. Apo receptor was generated from the atenolol complex by six washing steps of tenfold dilution in ligand-free buffer using a 1 h incubation period for the last three steps.

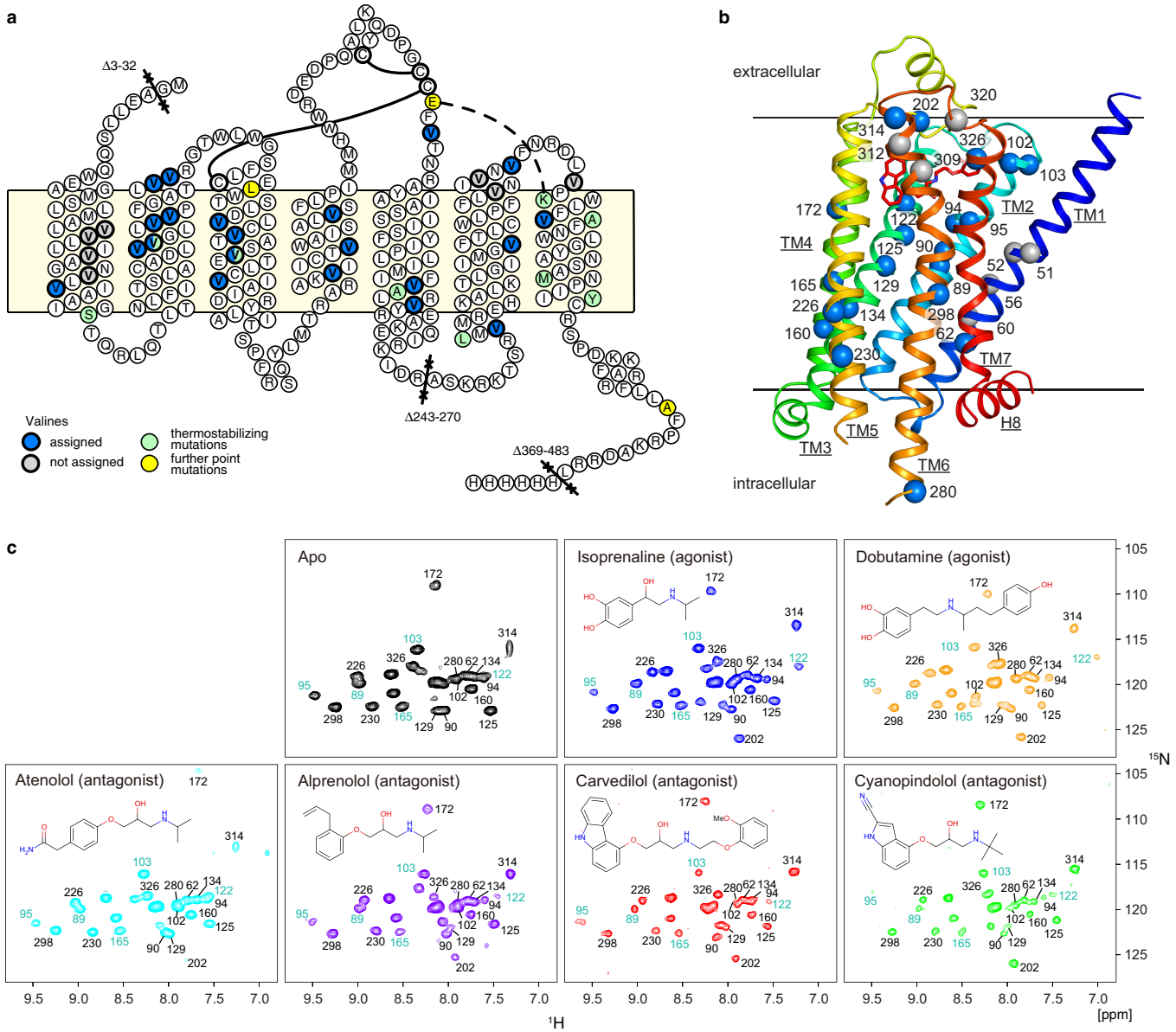
NMR NB80 binding experiment. Binding of NB80 to β_1AR mutants was assessed using TROSY and 1D proton NMR spectra. These spectra were recorded on the β_1AR mutants (TS-β_1AR: 132 μM, TS-β_1AR(A227Y): 120 μM, TS-β_1AR(L343Y): 110 μM, and TS-β_1AR(A227Y/L343Y): 120 μM) in the presence of saturating amounts (1 mM) of the agonist isoprenaline before and immediately after addition of an equimolar (relative to the receptor) amount of NB80. For TS-β_1AR(A227Y/L343Y) additional spectra were recorded after a further addition of the partial agonist cyanopindolol (1 mM) to the already present isoprenaline and NB80.

Scintillation proximity assay with ^{3}H-dihydroalprenolol. For pharmacological binding assays membranes were prepared from SF9 insect cells as described previously[26]. The total protein content of the membranes was estimated by A280 measurements using an average extinction coefficient of 1.0 per mg ml^{-1}. All assays were carried out in 96-well plates at 200 μg ml^{-1} total protein in membranes and 2 mg ml^{-1} WGA-YSi beads (Perkin-Elmer) in a 100 μl total volume per well. Samples were equilibrated at room temperature for at least 16 h. K_D values for the radioactive ligand [^{3}H]dihydroalprenolol (^{3}H-DHA) were determined by titrating ^{3}H-DHA from 0.032 to 100 nM. Non-specific binding was determined in presence of 1 μM S-propanolol to block the ligand binding site. Competition assays were performed in the presence of 20 nM ^{3}H-DHA (hot ligand) and increasing concentrations of the competitor (cold ligand). Dilutions of alprenolol, atenolol, cyanopindolol, dobutamine and isoprenaline were made with phosphate buffered saline (PBS, Biochrom, Germany). Due to the limited solubility of carvedilol in water, stock dilutions of carvedilol were prepared in DMSO. The final concentration of DMSO in the samples was 5%. To test the effect of NB80 binding on the affinities (IC_{50}) of isoprenaline for various receptor mutants, the competition assays were also carried out in the presence of saturating concentrations of NB80 (10 μM). IC_{50} values were determined by fitting the measured radioactive counts per minute CPM(X) at a specific concentration X of the competitor to the equation $\mathrm{CPM}(X) = \frac{\mathrm{CPM_{max}} - \mathrm{CPM_{min}}}{1 + \frac{X}{\mathrm{IC_{50}}}} + \mathrm{CPM_{min}}$, where $\mathrm{CPM_{max}}$ and $\mathrm{CPM_{min}}$ are maximal and minimal counts of the assay, respectively. The fits were carried in MATLAB (MathWorks, http://www.mathworks.com) with Monte-Carlo estimation of errors. K_i values were calculated from the obtained IC_{50} values according to the formula $K_i = \frac{\mathrm{IC_{50}}}{1 + \frac{A}{K_D}}$ where A is the concentration of the radioactive ligand and K_D is its affinity for the receptor determined in the direct binding experiment.

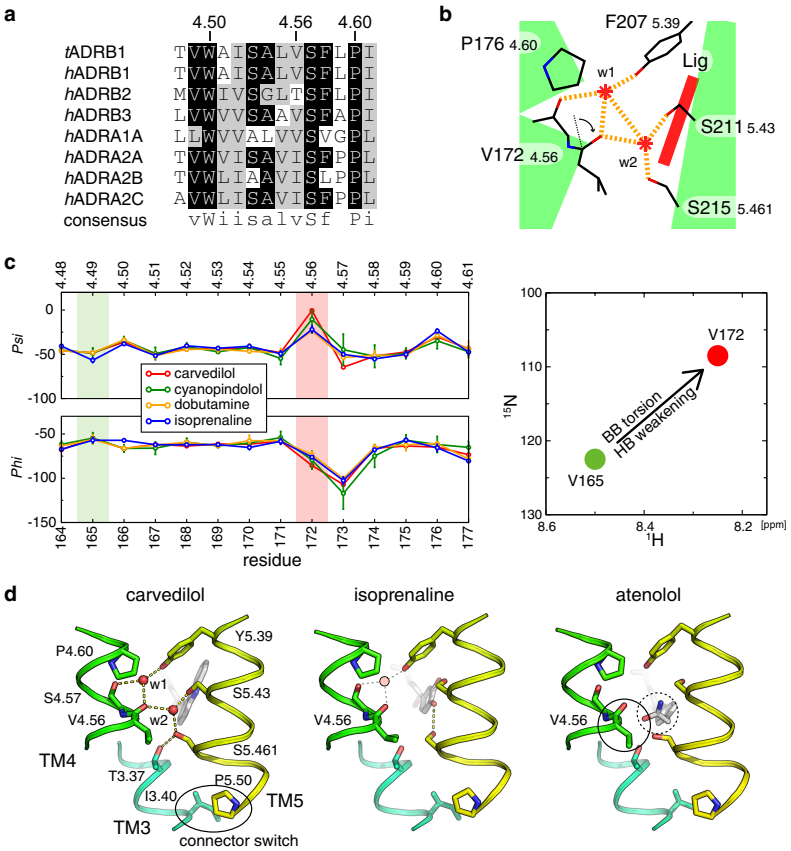
G protein activation assay. G protein activation was measured on purified β_1AR mutants reconstituted with MSP1E3D1[29] into POPC/POPG nanodiscs. MSP1E3D1 was expressed and purified as described[29] and cleaved with TEV protease. 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC, Avanti Polar Lipids) and 1-palmitoyl-2-oleoyl-*sn*-glycerol-3-phospho-(1′-*rac*-glycerol) sodium salt (POPG, Avanti Polar Lipids) were solubilized at a ratio of 1:1.5 (w/w) POPG/POPC in ND buffer (20 mM HEPES pH 8, 100 mM NaCl, 1 mM EDTA) with 50 mM sodium cholate (Sigma-Aldrich) at 4 °C. 133.3 μM MSP1E3D1 was incubated with 8 mM solubilized POPC/POPG and 10 μM purified β_1AR in ND buffer with a final concentration of 24 mM sodium cholate for 1 h at 4 °C. Nanodiscs containing the receptor were separated from empty nanodiscs using a cobalt-chelating resin. The heterotrimeric G protein was prepared by incubating 10 μM recombinant $G\alpha_{i1}$ and 10 μM native $G\beta\gamma_t$ in activation buffer (25 mM HEPES pH 7.5, 150 mM NaCl, 2 mM $MgCl_2$, 1 mM DTT) for 30 min at 4 °C.

G protein activation was detected by the change in tryptophan fluorescence caused by the exchange of GDP for GTPγS, associated conformational changes in the Gα subunit and its dissociation from the Gβγ subunit of the heterotrimeric G protein[30]. All measurements were carried out on a Varian Cary Eclipse fluorescence spectrophotometer (λ_{ex} = 295 nm, λ_{em} = 340 nm, 1.5 nm excitation slit, 20 nm emission slit, 2 s averaging time, 15 s cycle time) using final sample volumes of 1 ml in 10 × 4 mm cuvettes (Hellma, CH) and magnetic stirrers at 20 °C. Prior to activation, the fluorescence intensity baseline was recorded with 100 nM heterotrimeric G protein for approximately 500 s. The activation was started by adding 6 nM β_1AR and 10 μM GTPγS, and the fluorescence intensity was monitored for a further 1 h. For experiments in the presence of an agonist, the concentrated receptor stock solution (1.5 μM) was pre-incubated for 30 min at 4 °C with 40 μM isoprenaline, and the buffer during the measurements contained 2 μM of isoprenaline to maintain the saturation conditions for the receptor.

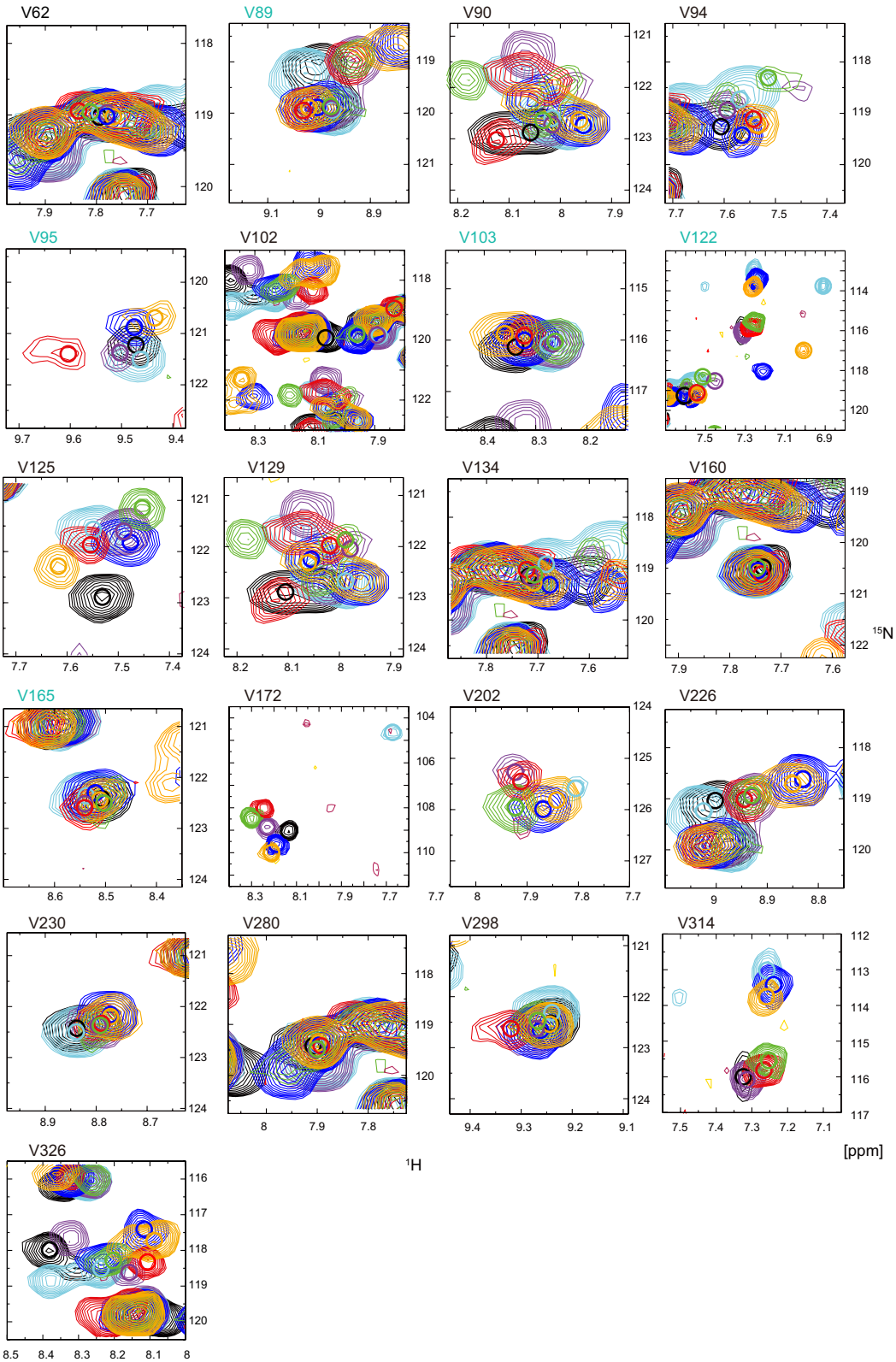
25. Warne, T. *et al.* The structural basis for agonist and partial agonist action on a β_1-adrenergic receptor. *Nature* **469,** 241–244 (2011).
26. Brueckner, F. *et al.* Structure of β-adrenergic receptors. *Methods Enzymol.* **520,** 117–151 (2013).
27. Alexandrov, A. I., Mileni, M., Chien, E. Y. T., Hanson, M. A. & Stevens, R. C. Microscale fluorescent thermal stability assay for membrane proteins. *Structure* **16,** 351–359 (2008).
28. Vajpai, N. *et al.* Backbone NMR resonance assignment of the Abelson kinase domain in complex with imatinib. *Biomol. NMR Assign.* **2,** 41–42 (2008).
29. Bayburt, T. H., Grinkova, Y. V. & Sligar, S. G. Self-assembly of discoidal phospholipid bilayer nanoparticles with membrane scaffold proteins. *Nano Lett.* **2,** 853–856 (2002).
30. Ernst, O. P., Bieri, C., Vogel, H. & Hofmann, K. P. Intrinsic biophysical monitors of transducin activation: fluorescence, UV-visible spectroscopy, light scattering, and evanescent field techniques. *Methods Enzymol.* **315,** 471–489 (2000).
31. Han, B., Liu, Y., Ginzinger, S. W. & Wishart, D. S. SHIFTX2: significantly improved protein chemical shift prediction. *J. Biomol. NMR* **50,** 43–57 (2011).



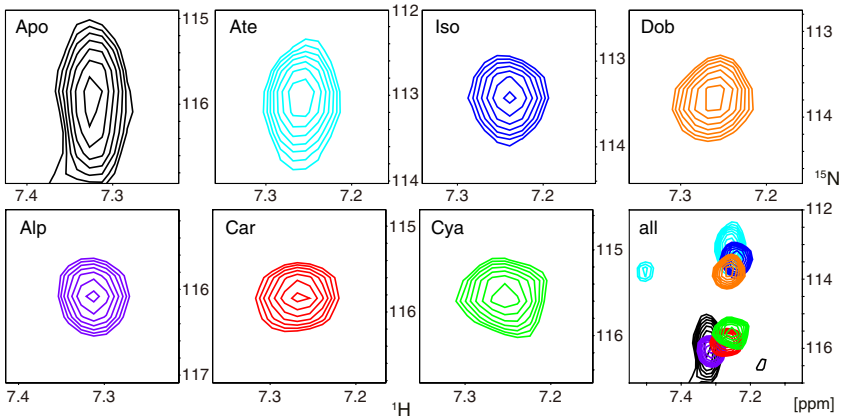
Extended Data Figure 1 | Position of valine residues in the thermostabilized mutant TS-β_1AR. a, Schematic representation of secondary structure and amino acid sequence of TS-β_1AR. As compared to the wild type, the TS-β_1AR has truncations at the N and C termini and the intracellular loop (ICL3) as well as nine thermostabilizing point mutations (colour coded green) and three additional point mutations (colour coded yellow). Valines labelled with ^{15}N are indicated by bold circles for assigned (blue) and unassigned (grey) residues. **b**, Structure of β_1AR in complex with carvedilol (PDB code 4AMJ). The protein backbone and carvedilol are shown in ribbon and red stick representation, respectively. The individual valines are depicted as spheres (blue, assigned; grey, not assigned) labelled by residue number. **c**, Full ^{1}H–^{15}N TROSY spectra of apo TS-β_1AR and all investigated ligand complexes. The ligand chemical structures are shown as inserts. Resonances are marked with assignment information (black, firm; cyan, tentative).



Extended Data Figure 2 | Effect of ligand head group substitution on hydrogen bond network involving V172(4.56) in β_1AR. a, Sequence alignment of turkey β_1AR and human adrenoreceptors around position 4.56 in TM4 showing the conserved proline at position 4.60. **b**, Schematic representation of the hydrogen bond network between TM4 and TM5 involving V172(4.56) in β_1AR. The hydrogen bond network (orange dotted lines) originates at the carbonyl of V172(4.56) and connects to TM5 via two water molecules (w1 and w2, red asterisks). **c**, Left, phi and psi dihedral angles (averages and standard deviations in case of multiple chains) in TM4 between residues 4.48 and 4.61 for different complexes of β_1AR (ligand, PDB code [chain]: carvedilol, 4AMJ [A,B]; cyanopindolol, 2VT4 [B,D] and 4BVN [A]; dobutamine, 2Y00 [A,B] and 2Y01 [A,B]; isoprenaline, 2Y03 [A,B]). V165(4.49) and V172(4.56), for which ^{1}H–^{15}N resonances could be observed, are shown on a green and red background, respectively. **c**, Right, average of the ^{1}H–^{15}N resonance positions for V165(4.49, green) and V172(4.56, red). The phi and psi values of V172(4.56) are distorted from the normal helical angles due to the loss of the intra-helical hydrogen bond to P176(4.60). Together with the loss of the canonical hydrogen bond, this strongly shifts both the ^{1}H and ^{15}N chemical shifts of V172(4.56) towards smaller p.p.m. values relative to V165(4.49), which has normal, helical phi and psi angles. **d**, Partial views of the crystal structures of the carvedilol complex (PDB code 4AMJ), the isoprenaline complex (PDB code 2Y03) and a docking model of the atenolol complex based on the cyanopindolol complex crystal structure (PDB code 4BVN). In the carvedilol complex, one water molecule (w2) forms a hydrogen bond network between the carbonyl oxygen of V172(4.56) and the side chains of S211(5.43) and S215(5.461). In the isoprenaline complex, the side chains of S211(5.43) and S215(5.461) are rotated and form hydrogen bonds to the catechol hydroxyl groups of isoprenaline. In this structure, no water molecule is observed at the equivalent position of w2. In the atenolol complex, the *para*-substituted acetamide of the ligand head ring (dashed circle) inserts between S211(5.43) and S215(5.461) and disturbs the interface between TM3 and TM5 near V172(4.56) (solid circle).

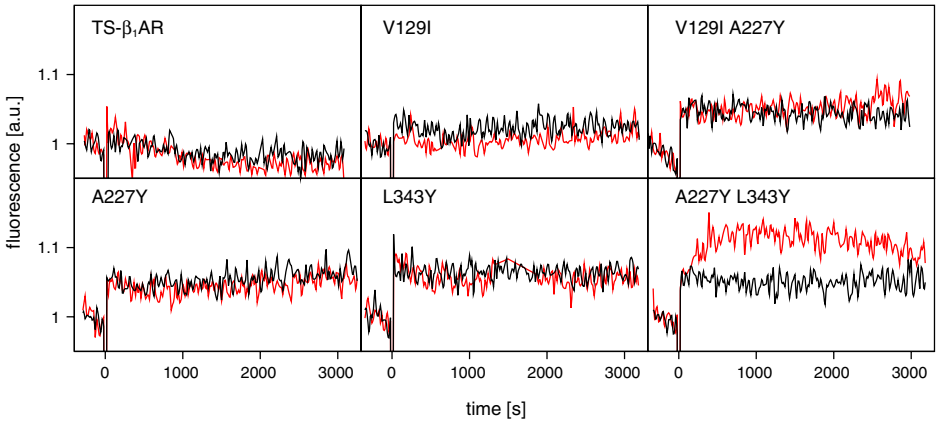


Extended Data Figure 3 | Response to various ligands for all assigned valine ^{1}H–^{15}N resonances in TS-β_1AR. Colour coding as in Extended Data Fig. 1: cyan (atenolol), blue (isoprenaline), orange (dobutamine), purple (alprenolol), red (carvedilol), green (cyanopindolol), black (apo receptor). For clarity the centres of resonances are marked by circles. Firmly (tentatively) assigned residues are marked in black (cyan).

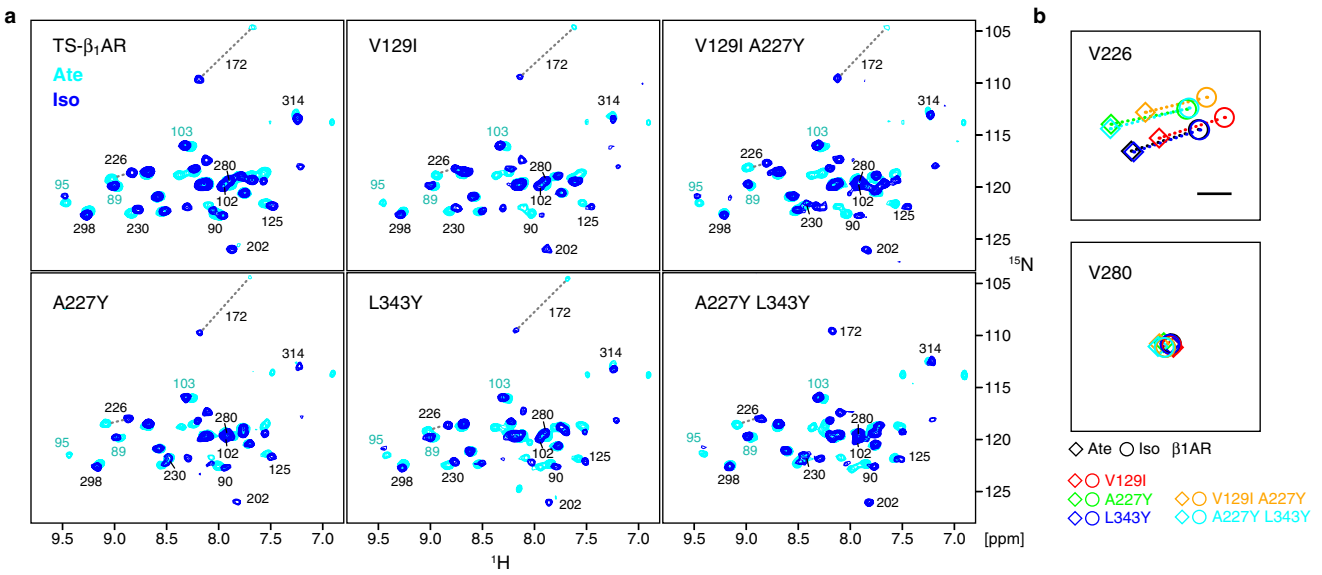


Ligand	Linewidth [Hz]	
	^{15}N	^{1}H
Apo	112.6	43.6
Ate	77.3	44.5
Iso	52.3	46.8
Dob	41.7	46.1
Alp	47.4	43.0
Car	40.0	50.4
Cya	37.7	48.1

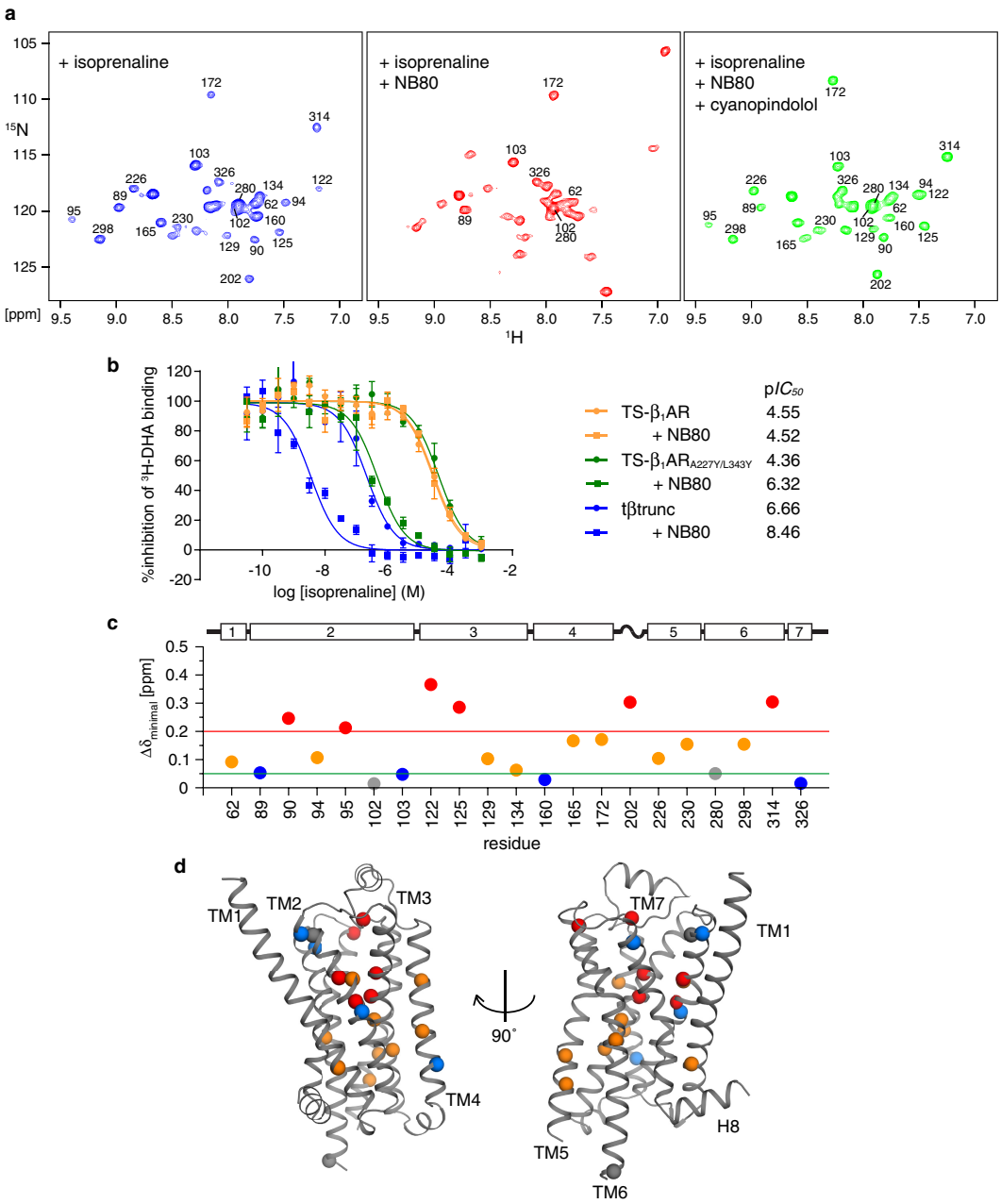
Extended Data Figure 4 | Evidence for micro- to millisecond dynamics at the ligand entry/exit pathway. Left, region of the ^{1}H–^{15}N TROSYs showing the V314(6.59) ^{1}H–^{15}N resonance of TS-β_1AR in the apo and various ligand-bound forms. The resonances of the apo and atenolol-bound forms are severely broadened in the ^{15}N dimension indicative of backbone dynamics in the micro- to millisecond range. The broadening is not observed for the other more tightly binding ligands. The resonances for the low affinity ligands, atenolol, isoprenaline and dobutamine, and for the high affinity ligands, alprenolol, carvedilol and cyanopindolol cluster at different positions. Interestingly, the resonance of the apo form clusters with the high affinity ligands, indicative of a similar backbone conformation. Right, ^{15}N and ^{1}H line widths of the V314(6.59) resonance for the apo and ligand-bound forms of TS-β_1AR.



Extended Data Figure 5 | Isoprenaline-induced G protein activation by β_1AR mutants. Activation was measured by the change in tryptophan fluorescence upon exchange of GDP for GTPγS in the Gα subunit of the heterotrimeric G_i protein. The figure shows the time courses of tryptophan fluorescence after the addition of different mutant forms of TS-β_1AR in nanodiscs and GTPγS to the heterotrimeric G protein in the presence (red) and absence (black) of the agonist isoprenaline. The increase in fluorescence intensity at time $t = 0$ is caused by the additional fluorescence of the receptor, the nanodisc scaffold protein and the ligand. For the active mutant TS-β_1AR(A227Y/L343Y), this initial rise is followed by an exponential fluorescence increase due to the activation and dissociation of the G protein. The rate of the $G\alpha_i$ activation by TS-β_1AR(A227Y/L343Y) (half-life ~290 s at 6 nM receptor) is approximately 4 times slower than for the activation by rhodopsin (half-life ~400 s at 1 nM receptor) using the same assay. This assumes that the K_m of $G\alpha_i$ for TS-β_1AR(A227Y/L343Y) is of the same scale as for rhodopsin (8.6 nM) and therefore significantly below the (saturating) concentration of $G\alpha_i$ (100 nM). Each mutant was measured three times except for TS-β_1AR(A227Y/L343Y) that was measured four times. Representative fluorescence curves from single experiments are shown.



Extended Data Figure 6 | Observed NMR effects of the reverse mutations of TS-β_1AR towards the native β_1AR sequence. a, ^{1}H–^{15}N TROSY spectra of TS-β_1AR and several reverse single and double mutants in complex with either atenolol (cyan) or isoprenaline (dark blue). Resonances are marked with assignment information (black, definite; cyan, tentative). **b**, Enlarged regions of the ^{1}H–^{15}N correlation spectra showing only the resonance positions of V226(5.57) and V280(6.25) for all mutants in both ligand-receptor complexes. The black bar represents 0.1 p.p.m. in ^{1}H and 1 p.p.m. in ^{15}N. For all mutants, the resonances for V226(5.57) show efficacy-related chemical shift changes between atenolol (diamonds) and isoprenaline (circles). For the A227Y mutants TS-β_1AR(A227Y) (green), TS-β_1AR(V129I/A227Y) (orange) and TS-β_1AR(A227Y/L343Y) (cyan), the ^{1}H–^{15}N resonances of V226(5.57) also exhibit an overall shift due to a ring current effect from the introduced Y227(5.58) side chain, which has no structural significance. However, the V129I mutants TS-β_1AR(V129I) (red) and TS-β_1AR(V129I/A227Y) (orange) exhibit further ^{1}H–^{15}N shifts towards a more active (that is, bent) state of TM5 relative to the mutants that carry the V129(3.40) residue. For all mutants, the V280(6.25) resonances fall basically in identical positions and show no changes between atenolol and isoprenaline.



Extended Data Figure 7 | Binding of NB80 to agonist-activated TS-β_1AR(A227Y/L343Y). a, Sequential addition of isoprenaline (1 mM, blue), NB80 (120 μM, red), and cyanopindolol (1 mM, green) to TS-β_1AR(A227Y/L343Y) (120 μM) followed by ^{1}H–^{15}N TROSY spectra. Addition of an equimolar amount of NB80 changes the spectrum drastically. Except for few resonances (62, 89, 102, 103, 172, 280 and 326), the assignments are not transferable from the isoprenaline complex. The addition of cyanopindolol to this mixture apparently replaces isoprenaline in the receptor and drives the spectrum to the purely cyanopindolol-bound form. These results indicate that TS-β_1AR(A227Y/L343Y) can be activated, but that the fully active state requires stabilization with a G protein-like partner at the intracellular site. **b**, Changes in isoprenaline affinity of β_1AR induced by NB80 binding. The isoprenaline affinity was assayed for TS-β_1AR (orange), TS-β_1AR(A227Y/L343Y) (green), and the truncated native turkey β_1AR receptor (tβtrunc, blue)[17] by radio-ligand competition using ^{3}H-dihydroalprenolol in the presence of saturating amounts of NB80 (10 μM, squares) or its absence (circles). For TS-β_1AR(A227Y/L343Y) and tβtrunc the apparent affinity increases by two log units in the presence of NB80, whereas no change occurs for TS-β_1AR. Data points are shown as mean and standard deviation of three independent experiments for each condition. **c**, Combined ^{1}H and ^{15}N chemical shift changes ($\Delta\delta = (\Delta\delta_{1H}^2/2 + \Delta\delta_{15N}^2/50)^{1/2}$) between the isoprenaline-TS-β_1AR(A227Y/L343Y) and the NB80-isoprenaline-TS-β_1AR(A227Y/L343Y) complexes as a function of residue number. Except for the few assigned peaks in the NB80-isoprenaline-TS-β_1AR(A227Y/L343Y) complex, chemical shift deviations are calculated as the minimal difference from the isoprenaline peak position to the nearest peak position in the isoprenaline–NB80 complex. Note that these minimal chemical shift deviations present a lower boundary for the deviations that could be obtained from a full assignment of the isoprenaline–NB80 complex. Peak deviations are colour-coded in red (≥0.2 p.p.m.), orange (0.2–0.05 p.p.m.), blue (<0.05 p.p.m.), and grey (ambiguity from peak overlap). **d**, Representation of the chemical shift response to nanobody binding given in **c** on the structure of β_1AR. Valines in TM1 and TM7 do not show a large chemical shift response. Due to overlap the response for V280(6.25) in TM6 is unclear. Valines in TM2 show a strong response around the sodium binding region near residue D87(2.50), but valines at its extracellular side are not responsive. In general, valine resonances at the extracellular sides of TM3–TM6 are severely shifted. This indicates that the NB80 binding at the intracellular side has a long-range effect on the extracellular side near the ligand binding site.

Extended Data Table 1 | Pharmacological and geometrical properties of the β_1AR ligands used in this study

	pharmacology				geometry	
	pK_D β36 -m23*	pK_i TS-β_1AR†	pIC_{50} TS-β_1AR‡	Gs efficacy *t*βtrunc§	insertion depth‖	tail volume¶
cyanopindolol	10.01 ± 0.11	>8.4 ± 0.3#	>8.1 ± 0.1#	39 ± 1	7.22	90.11
carvedilol	8.72 ± 0.09	>8.3 ± 0.3#	>7.9 ± 0.1#	12 ± 0.4	7.62	163.47
alprenolol	7.35 ± 0.07	7.6 ± 0.3	7.2 ± 0.1	31 ± 2	7.25	73.87
dobutamine	5.49 ± 0.03	5.0 ± 0.3	4.6 ± 0.1	112 ± 7	8.46	170.34
isoprenaline	5.16 ± 0.05	4.9 ± 0.3	4.6 ± 0.2	100	7.92	73.87
atenolol	4.29 ± 0.11	5.0 ± 0.3	4.6 ± 0.2	2.1 ± 1	ND⋆	73.87

*Values for binding affinity correspond to the β36-m23 construct, which contains C- and N-terminal deletions and six thermostabilizing mutations[10]. The $pK_D = -\log K_D$ [M] was obtained from [^{3}H] CGP12177 whole-cell binding assays in stably expressed CHO cells[17].

†pK_i values were obtained from the pIC_{50} values given in column 3 by adding $\log(1+[lig]/K_D) = 0.366$ as a correction for the concentration of the [^{3}H]dihydroalprenolol (^{3}H-DHA) ([lig] = 20 nM) and its determined dissociation constant ($K_D = 15.1 \pm 7$ nM, see Methods). Errors are obtained by error propagation from the errors of the pIC_{50} values and K_D.

‡pIC_{50} values were determined by the radioligand inhibition assay described in the Methods section using ^{3}H-DHA. Data are given as mean and standard deviation of three independent experiments.

§Values for efficacy towards Gs-mediated signalling as determined by adenylyl cyclase activation of the truncated native turkey β_1AR receptor (*t*βtrunc)[17]. The efficacies are indicated as the percentage of the (maximal) isoprenaline response obtained from ^{3}H-cAMP accumulation assays in CHO cells.

‖The insertion depth (Å) of the ligand was taken as the distance between the β-carbon atom of the ligand amino group and the amide nitrogen atom of V125 (V117 for β_2AR) in the crystal structures of turkey β_1AR in complexes with isoprenaline (PDB ID: 2Y03), dobutamine (PDB ID: 2Y00), carvedilol (PDB ID: 4AMJ), and cyanopindolol (PDB ID: 4BVN) as well as of human β_2AR in complex with alprenolol (PDB ID: 3NYA).

¶The tail volumes (Å^3) were calculated by the Molinspiration Property Calculation Service (http://www.molinspiration.com/cgi-bin/properties) for the tail group including the amino moiety.

#The pK_i and pIC_{50} are limited by the K_D and concentration of the radioactive tracer ligand as well as by the concentration of the receptor in the competition assay.

⋆Not determined, no crystal structure available.

Extended Data Table 2 | Sequence-specific assignment $^{1}H–^{15}N$ valine resonances

Assigned residue	Position	Point mutation	Spectra with ligands	Further information
V62	1.53	V62A	Ate, Apo, Alp	
V90	2.53	V90A	Ate, Alp, Car, Cya	HN(CO)
V94	2.56	V94A	Alp, Car, Cya	
V102	2.64	V102A	Ate, Apo, Alp, Iso, Dob, Car, Cya	
V122	3.33	V122A	Alp, Car, Cya	
V125	3.36	V125A/I	Alp(A), Car(A), Ate(I), Apo(I)	
V129	3.40	V129I	Ate	
V134	3.45	V134A	Ate	
V160	4.44	V160A	Ate	
V172	4.56	V172A	Ate, Alp	Pred
V202	ECL	V202A	Alp	Pred
V226	5.57	V226A	Alp, Car, Cya	
V230	5.61	V230A	Alp	
V280	6.25	V280A	Alp, Car, Cya	
V298	6.43	V298A	Ate	HN(CO)
V314	6.59	V314A	Alp, Car, Cya	
V326	7.36	V326A	Alp	
V89	2.52	-	Ate, Apo, Alp, Iso, Dob, Car, Cya	HB, NM
V95	2.57	-	Ate, Apo, Alp, Iso, Dob, Car, Cya	HB, NM, HN(CO)
V103	2.65	-	Ate, Apo, Alp, Iso, Dob, Car, Cya	HB, NM, HN(CO)
V165	4.49	-	Ate, Apo, Alp, Iso, Dob, Car, Cya	NM

Ate, atenolol; Alp, alprenolol; Iso, isoprenaline; Dob, dobutamine; Car, carvedilol; Cya, cyanopindolol
HN(CO): information from 2D-filtered HN(CO) of [^{15}N,^{13}C]Gly,Val-, [^{15}N]Leu-labelled receptor
Pred: ShiftX2 ^{15}N chemical shift prediction[31]
HB: expected $^{1}H^{N}$ shift based on hydrogen bond length analysis
NM: peak shift due to mutation of neighbouring residue
Assignments for V89(2.52), V95(2.57), V103(2.65), V122(3.33), and V165(4.49) are tentative. V51(1.42), V52(1.43), V56(1.47), V60(1.51), V309(6.54), V312(6.57), and V320(ECL) have not been assigned.

Extended Data Table 3 | Summary of the NMR response to ligands, G protein activation and NB80 binding data for the original TS-β_1AR construct and various reverse mutants towards the native β_1AR sequence

mutant name	residue			V226 response*	V280 response*	G protein activation†	NB80 binding‡	T_m [˚C]§
	129	227	343					
original TS-β_1AR	V	A	L	+	-	-	- (4.55/4.52)	58.9 +/– 0.6
TS-β_1AR$_{V129I}$	**I**‖	A	L	+	-	-	N.D.¶	58.6 +/– 0.2
TS-β_1AR$_{A227Y}$	V	**Y**	L	+	-	-	-	54.9 +/– 0.1
TS-β_1AR$_{L343Y}$	V	A	**Y**	+	-	-	-	54.7 +/– 0.3
TS-β_1AR$_{V129I/A227Y}$	**I**	**Y**	L	+	-	-	N.D.	54.4 +/– 0.1
TS-β_1AR$_{A227Y/L343Y}$	V	**Y**	**Y**	+	-	+	+ (4.36/6.32)	47.8 +/– 0.2

*Determined from the shifts of the ^{1}H–^{15}N resonances in response to binding of atenolol and isoprenaline (see Extended Data Fig. 6B).
†Determined from the G_i activation assay described in Extended Data Fig. 5.
‡Determined from NMR experiments by mixing of isoprenaline-activated TS-β_1AR with NB80. The ^{1}H–^{15}N spectra are shown for TS-β_1AR(A227Y/L343Y) in Extended Data Fig. 7. The spectra of the other TS-β_1AR mutants showed no response to addition of NB80 up to equimolar concentrations of ~100 μM for both proteins. Values in parentheses show pIC_{50} for isoprenaline binding in the absence or presence of NB80, respectively (see Extended Data Fig. 7).
§Melting temperature T_m of the apo form receptor determined by the CPM thermal shift assay.
‖Native residues are highlighted in bold.
¶Not determined.

3.3 A NOE-based approach to backbone resonance assignment of the β_1-adrenergic receptor

Introduction

Sequential assignments of the protein backbone are commonly obtained by triple resonance correlation experiments such as the HNCA and the HNCOCA (*119*, *120*). These experiments involve magnetization transfer via one-bond and two-bond scalar coupling interactions. Whereas this approach is very successful for small- to medium-size proteins and has seen various technical improvements such as transverse relaxation optimized spectroscopy (TROSY) (*121*), it is of limited use for larger proteins (>100 kDa). Slow molecular motion of these large proteins in solution increases transverse relaxation rates, which decrease the efficiency of transfer and frequency labeling steps in triple resonance experiments.

An alternative strategy for sequential assignments may be based on ^{15}N-edited nuclear Overhauser enhancement spectroscopy (NOESY) to obtain spatial inter-proton information (*122*, *123*) resolved for the backbone amide nitrogen. This approach allows correlating amide protons of neighboring residues within 5 Å distance and thus may be applied for sequential assignment of the amide backbone.

This section describes uniform labeling by ^{15}N and deuteration of the TS-β_1AR in insect cells, which was achieved by growth on ^{2}H,^{15}N-labeled yeast extract. This approach has allowed obtaining ^{15}N-edited 3D NOESY spectra of very high quality. An initial assignment is reported for a selected stretch of residues in the transmembrane helix 5, which is part of the allosteric signaling network of the receptor (*103*). Finally, the impact of fractional deuteration for extended assignment of the TS-β_1AR is discussed.

Results and Discussion

Uniform isotope labeling of β_1-adrenergic receptor in insect cells

TS-β_1AR was expressed in Sf9 insect cells using 8 g/L ^{2}H,^{15}N-labeled YE and 1 g/L $^{15}N_2$-L-glutamine yielding 1 mg/L of purified receptor. The expression level of ^{2}H,^{15}N-TS-β_1AR was comparable to unlabeled production of the receptor in commercial medium. Moreover, the fraction of active receptor obtained by affinity chromatography (see Materials and Methods) was similar for unlabeled TS-β_1AR and ^{2}H,^{15}N-TS-β_1AR. Therefore, the ability to bind ligands and the yield remained unchanged when homemade YE was used for labeled expression of TS-β_1AR. A detailed analysis of ^{2}H,^{15}N-TS-β_1AR by NMR is described in Chapter 2.1.

Sequential assignment of backbone amide resonances

^{15}N-edited 3D NOESY spectra using a TROSY-type pulse sequence (*124*) were recorded on the 110 µM ^{2}H,^{15}N-TS-β_1AR solubilized in a *n*-decyl-β-D-maltopyranoside detergent micelle. In total, 243 backbone amide resonances were observed in the diagonal of the 3D NOESY, which correspond to 80 % of the expected 303 residues excluding the N-terminus and 11 proline residues. Moreover, 474 cross peaks between these amide resonances were

obtained. These include 138 H^N_i-N^H_j / H^N_j-N^H_i mirror image pairs. Moreover, long-range *i-2* and *i+2* connectivity was observed for A227 and V230 (Figure 3.4A).

Figure 3.4A shows sequential strip plots obtained from the 3D experiment for the receptor in complex with alprenolol. The assignment process was based on the assigned amide resonances for the ^{15}N-valine labeled receptors (Chapter 3.2). A consecutive sequence of six residues is depicted including V226 and V230 in the transmembrane (TM) helix 5. Figure 3.3B indicates where these residues are located within the receptor. TM5 and TM6 as well as their intervening loop ICL3 are colored from orange to red. These structural elements of TS-β_1AR mediate G protein binding. All assigned residues are contained in TM5, which is part of the allosteric activation pathway and displayed a remarkable unified response to binding of different ligands (Chapter 3.2). So far, chemical shift changes obtained for the ^{1}H-^{15}N resonances of V226 alone were used to monitor this response. Therefore, it is expected that the assigned sequence of consecutive residues may broaden the available information for binding of distinct ligands to TS-β_1AR.

Furthermore, 315 cross peaks were observed in the range from -1 to 5 ppm of the indirect ^{1}H dimension indicating only partial deuteration. This is e.g. evident in the methyl region of the 3D NOESY for the assigned residues F225 to V230 (Figure 3.4A, upper panel). The obtained peaks mainly result from intra-residue NOEs. However, F225 clearly shows a strong cross peak at 0.84 ppm, which is not expected for a phenylalanine side chain resonance (*125*). Consequently, this cross peak corresponds to an inter-residue NOE from a neighboring residue. In particular, the resonance very likely originates from the δ-CH_3 methyl group of the adjacent I224.

The observed partial deuteration (Figure 3.4A) can be explained by two reasons: (*i*) incomplete deuteration of the prepared YE and (*ii*) an unlabeled pool of amino acids as well as metabolic scrambling of non-essential amino acids in insect cells. As shown by mass spectroscopic analysis of individual amino acids, an average deuteration level of 75 % was achieved based on growth of *Pichia* in D_2O using protonated glucose (Chapter 2.1). This relates to the biosynthesis of amino acids involving glucose as a precursor. Insect cells share some of these metabolic pathways, which result in dilution of the deuterium label for several non-essential amino acids such as alanine and glycine (Chapter 1). Further, the initial starvation step of the labeling protocol (Chapter 2.1) may not completely deplete the pool of unlabeled amino acid in insect cells. Hence, these amino acids are likely to be incorporated into TS-β_1AR during expression.

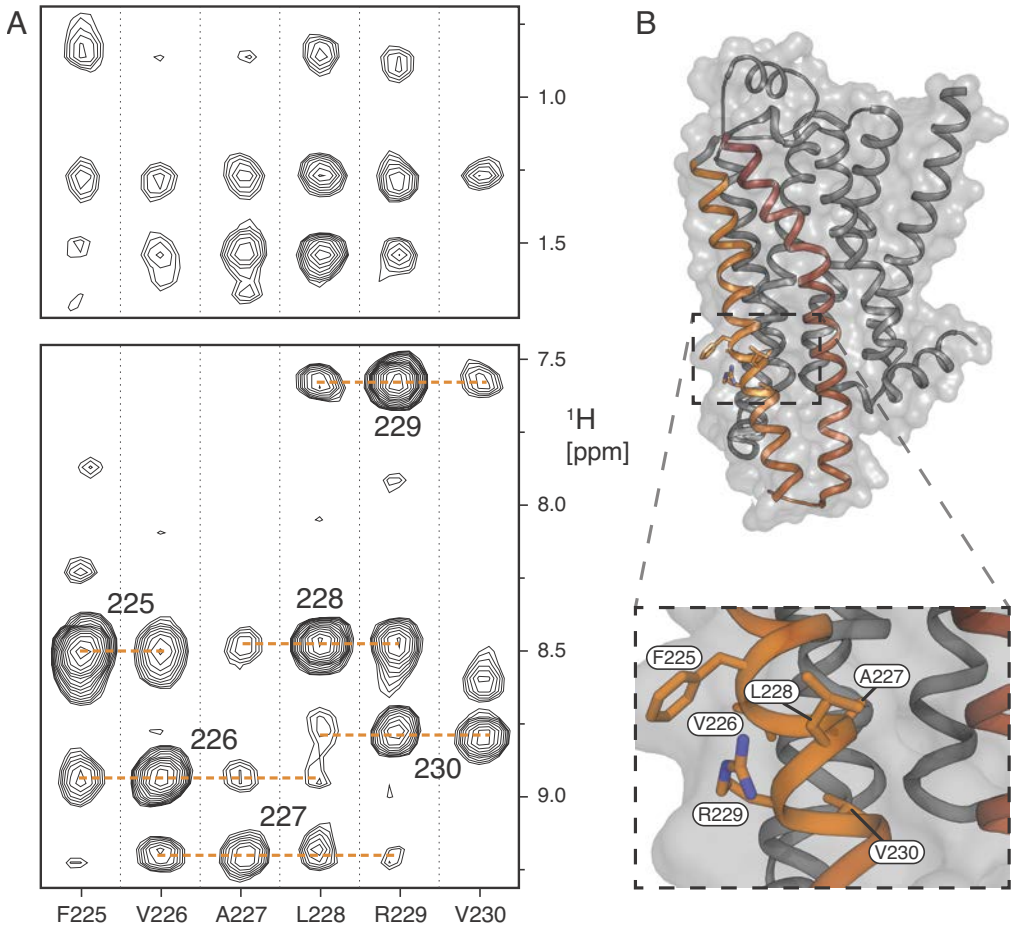


Figure 3.4 Sequential backbone assignment of TS-β_1AR based on a ^{15}N-edited 3D NOESY-TROSY-HSQC experiment recorded on ^{2}H,^{15}N-labeled receptor in complex with alprenolol. **A** Strip plots showing a consecutive stretch of amino acids. *Bottom*: Amide region indicating the sequential walk used to assign the amide backbone resonances (orange dashed line). *Top*: Selected methyl region containing aliphatic cross peaks. V226 and V230 have been assigned using site-directed mutagenesis in the selectively labeled receptor (*103*). ^{2}H,^{15}N-TS-β_1AR was expressed in Sf9 cells grown on ^{2}H,^{15}N-yeast extract and $^{15}N_2$-glutamine. The ^{15}N-edited 3D NOESY-TROSY-HSQC experiment was recorded on 110 µM of labeled receptor for 10 days. **B** Localization of the assigned amino acids in the backbone of ultra-stable β_1AR (PDB entry 4BVN). The transmembrane helices TM5 and TM6 as well as the intervening loop ICL3 are colored using a gradient from orange to red. A detailed view of TM5 is shown for F225 to V230 (dashed box).

Residual protonation of the side chain positions may lower the sensitivity of the ^{15}N-edited 3D NOESY experiment. However, fractional deuteration is to some extent advantageous to derive complementary information for the assignment process (*126*). Moreover, the specific interactions of the receptor with its ligands involve side chain contributions, which have been used recently to study dynamics of selectively labeled GPCRs by NMR (*41*, *107*).

Conclusions and perspectives

Uniform isotope labeling by ^{15}N and fractional deuteration of TS-β_1AR was achieved in insect cells using labeled yeast extract. This approach enabled high-resolution NMR studies

on the labeled receptor. Well-resolved ^{15}N-edited 3D NOESY spectra were obtained and thereby a preliminary assignment of a stretch of residues ranging from F225 to V230. Based on the obtained 138 pairs of mirror peaks, it is expected that the backbone resonance assignment of TS-β_1AR can be further advanced. Moreover, labeling by ^{13}C may complement the available spectral information and can be achieved by preparation of the respective ^{2}H,^{13}C-labeled yeast extract.

Assignments obtained for TS-β_1AR may be transferred to constructs for which thermostabilizing mutations have been reverted towards the wild type of the receptor and thus displayed the ability to bind the Nb80 nanobody (Chapter 3.2). The combination of fractional deuteration and selective amino acid labels may also be used to resolve spectral overlap in 2D ^{1}H,^{15}N-TROSY experiments. A similar analysis as shown for TS-β_1AR (*103*) may be possible for the β_2-adrenergic receptor, for which so far only analysis of selectively labeled [^{13}C-*methyl*]methionine residues has been carried out (*29*).

Fractional deuteration and uniform labeling by ^{15}N and ^{13}C of GPCRs in insect cells may further provide a basis for solid-state NMR (ssNMR) studies. Recently, a structure of the chemokine receptor CXCR1 expressed in *E. coli* was determined by ssNMR (*108*). This strategy was based on refolding of the receptor, which is often challenging and may not yield an active protein. In contrast, most GPCRs expressed in insect cells were obtained in their active form and thus allow to study ligand interaction and possibly signaling of the functional receptor (*127*).

Materials and Methods

Uniform isotope labeling in insect cells

Detailed protocols are presented in Chapter 2.1.

TS-β_1AR expression and purification

Detailed protocols are presented in Chapter 3.2.

NMR sample preparation

The NMR sample was prepared in a Shigemi tube as a volume of 270 µl containing 110 µM of ^{2}H,^{15}N-TS-β_1AR in 20 mM TRIS (pH 7.5), 1 mM of alprenolol, 100 mM NaCl, 0.1 % DM, and 5 % D_2O.

NMR spectroscopy

All NMR experiments on ^{2}H,^{15}N-TS-β_1AR were recorded on a Bruker Avance III 900 MHz spectrometer equipped with a triple-resonance cryoprobe at a temperature of 30 °C. The ^{15}N-edited 3D NOESY-TROSY experiment was recorded with 200 (^{1}H) x 160 (^{15}N) x 1024 (^{1}H) complex points, a mixing time of 55 ms, acquisition times of 9.6 ms (^{1}H) x 25.6 ms (^{15}N) x 43 (^{1}H) ms, and a total experimental time of 10 days. INEPT transfer times were optimized for amide proton T_2 values of ^{2}H,^{15}N-TS-β_1AR (3 ms).

NMR data analysis and assignment

NMR data were processed with the NMRPipe software package (*128*). Assignments of 3D NOESY spectra were carried out using the program SPARKY (*129*).

4 Combined use of yeast and algal cell extracts for uniform isotope labeling of the human lipopolysaccharide-binding protein in insect cells

Introduction

Recombinant baculovirus-mediated expression in insect cells allows producing a multitude of human proteins in sufficient quantities for a variety of structural studies. The system is frequently applied for the production of cytosolic proteins. However, cytosolic expression is not possible when the protein expressed shows toxicity or undesirable interactions within the host cell. Therefore, secreted expression may be the only available option in spite of the fact that generally lower amounts of protein are obtained. For complex secreted proteins, deterioration of the secretory pathway by the progressing baculovirus infection leads to inefficient secretion and a decrease in yield (*130*). Further engineering of the secretory pathway was shown to improve expression yields of secreted proteins in the insect cell system (*131*).

Recently, its was shown that the secretion pathway in the baculovirus/insect cell system led to production of a soluble form of the murine lipopolysaccharide-binding protein (mLBP) suitable for crystallization studies (*132*). The solved structure of mLBP shows different glycosylation sites as seen in Figure 4.1. Glycosylation is required to enter the secretory pathway which enables secretion of glycoproteins (*133*). While mLBP was characterized in detail by X-ray crystallography, its interaction with the lipopolysaccharide (LPS) remains unresolved at atomic resolution. Nuclear magnetic resonance (NMR) may be well suited for such interaction studies, but poses the additional challenge of stable isotope labeling of the protein. Only recently substantial methodological progress has been made that enables such uniform isotope labeling including deuteration in insect cells in an economical manner (*39*, *40*, *98*).

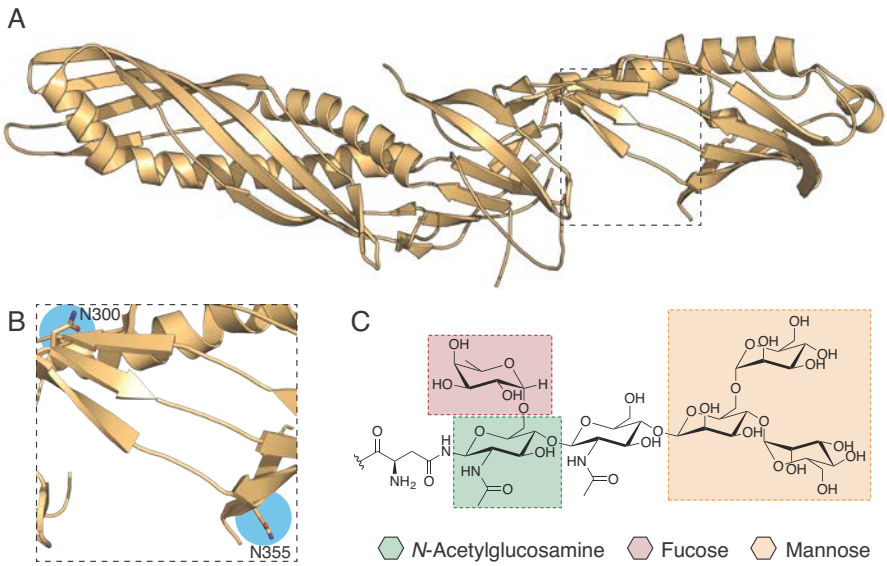


Figure 4.1 LBP and glycosylation in insect cells. **A** Structure of murine LBP (PDB entry 4M4D) determined by X-ray crystallography. **B** A detailed view of the two reported glycosylation sites. Asparagine residues N300 and N355 (blue circles) are *N*-glycosylated during post-translational modification to facilitate secretion of mLBP. **C** Paucimannosidic *N*-glycan pattern. The paucimannose-type structure is the most abundant *N*-glycosylation pattern obtained in insect cells (*134*).

This chapter describes uniform ^{15}N-labeling of the secreted hLBP by combining commercial algal and homemade yeast extracts. While labeling solely based on yeast extract reduced the amounts of secreted protein, supplementation of algal extract restored expression of hLBP. Also, labeling by algal extract alone resulted in high-level expression of hLBP, but still required addition of tryptophan. The amino acid analysis of both extracts indicated that the difference in expression level was not related to the amino acid composition, but must be rather due to other factors such as differences in the carbohydrate content. Finally, a well-resolved ^{1}H-^{15}N-TROSY-HSQC of ^{15}N-hLBP was obtained on a sample produced on algal and yeast extracts at a 1:1 ratio.

Results and Discussion

Development of a labeling strategy for secreted expression of LBP in insect cells

Secreted expression of hLBP in *Sf9* insect cells was achieved by a recombinant N-terminal fusion to the honeybee melittin signal sequence (MKFLVNVALVFMVVYISYIYAD-hLBP(A^{26}-V^{481})-GSGSGSHHHHHHHHHH). The secreted protein was purified from the medium with immobilized-metal affinity chromatography which enables robust quantification of the protein yield under different labeling conditions. Initially, reference expression levels were established in commercial SF-4 and Lonza media (Figure 4.2). Remarkably, 2.4 mg/L of purified hLBP were obtained in SF-4 (Figure 4.2, condition 1), while expression in Lonza yielded only 1.2 mg/L (Figure 4.2, condition 2). This twofold reduction in yield can be rationalized by the prior adaptation of the Sf9 cells in SF-4, which is solely based on YE (personal communication with suppliers' scientific support). In contrast, Lonza Insect Xpress medium contains two different undisclosed sources of hydrolysates (personal communication with suppliers' scientific support) and SF-4 adapted insect cells appear to perform less well in this medium. Thus, all subsequent labeling schemes were tested in SF-4 adapted insect cells.

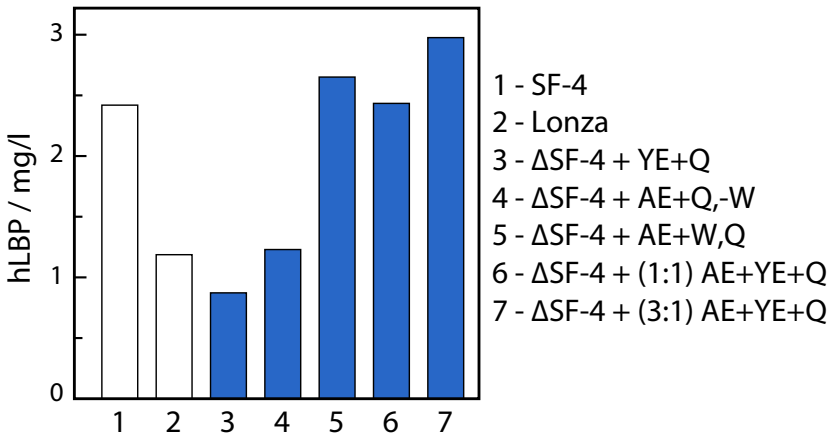


Figure 4.2 Expression yield of hLBP Sf9 insect cells under different conditions. Reference expression was established in commercial SF-4 medium (1). Additionally, expression in commercial Insect Xpress medium (2) following adaptation to SF-4 medium was carried out. Different supplements to SF4 medium depleted in amino acids and yeastolate (ΔSF4) were tested as follows: 8 g/L of yeast extract (YE) + 1 g/L Q (3), 8 g/L of algal extract (AE) + 1 g/L Q without W (4), and 8 g/L of AE + 1 g/L Q and 20 mg/L W (5). Also, two ratios of AE and YE supplemented with 1 g/L of Q were tested using 4 g/L of AE + 4 g/L of YE (6) and 6 g/L of AE + 2 g/L of YE (7), respectively. Sf9 cell cultures were handeled as described in Chapter 2.1.

Since a uniform ^{15}N-labeling scheme based on yeast extract was previously successful for expression of soluble and membrane protein in insect cells (*98*), the same scheme was employed for expression of hLBP. Hence, an amino acid-depleted, yeast extract-free SF-4 medium (ΔSF4) was supplemented with 1 g/L of glutamine and 8 g/L of homemade yeast extract (YE) and the expression of hLBP was analyzed (Figure 4.2, condition 3). However, only 0.9 mg/L of purified hLBP were obtained which amounts to 36 % of the reference expression level in full SF-4 medium. This significant reduction had not been observed for the previously expressed proteins.

To further analyze the cause for the reduction in protein yield, a commercial algal lysate based extract (ISOGRO, Sigma) was used as an alternative source of amino acids in ΔSF4. The feasibility of this approach has been demonstrated recently (*39*). In contrast to YE, algal extract (AE) is the result of hydrolysis of algal cells by mineral acids. Expression of hLBP in ΔSF4 supplemented with 8 g/L of AE and 1 g/L glutamine yielded 1.2 mg/L of hLBP, which is an increase in the expression level by 33 % over YE (Figure 4.2, condition 4). However, the amino acids cysteine and tryptophan are lost during conventional acid hydrolysis for the preparation of commercial AE, while glutamine and asparagine are converted to glutamic acid and aspartic acid, respectively (*135*). In line with the previously published labeling protocol based on AE (*39*), restoration of the expression yield to 2.6 mg/L of purified hLBP was obtained by supplementing 1 g/L glutamine and 20 mg/L of tryptophan to the expression medium (Figure 4.2, condition 5). Moreover, supplementation of cysteine was not required for recovery of the expression yield as has been shown for the YE based labeling scheme (Chapter 2.1). Biosynthesis of cysteine from methionine in insect cells was previously reported and may explain this result (*36*).

In the next step, the prospect of a labeling strategy based on the combination of both extracts was analyzed. Since YE is an autolysis-based extract, it still contains tryptophan at a growth-promoting level for the insect cells and thus has the potential to replace the additional amino acid supplement needed for AE. Combining the two extracts may indicate whether a specific component may be absent in the homemade YE but not in the commercial AE. Furthermore, YE may contain an inhibitory factor related to the secretion of hLBP. Therefore, ΔSF4 was supplemented with 1 g/L of glutamine and two different ratios of AE/YE accounting for a total extract concentration of 8 g/L without addition of tryptophan. Supplementing ΔSF4 with 4 g/L of AE and 4 g/L of YE (ratio 1:1) yielded 2.4 mg/L of purified hLBP (Figure 4.2, condition 6). When 6 g/L of AE and 2 g/L of YE (ratio 3:1) were used, the protein yield was again increased by 25 % to 3 mg/L (Figure 43.2, condition 7).

Interestingly, AE supplemented with tryptophan and the combined AE/YE were able to overcome the low expression yield observed for homemade YE provided that glutamine was supplemented for all of the conditions. These preliminary data suggest one or more limiting components in the homemade YE as the probable cause for the observed recovery of the protein yield upon blending of both extracts. Conversely, an inhibitiory effect of homemade YE on hLBP expression was not observed for concentrations up to 4 g/L.

Amino acid composition of algal extract

To further characterize the commercial AE, the amino acid content was quantified by phenylthiohydantoin amino acid analysis (PTH-AAA). The sample preparation and data analysis was performed according to the previously established PTH-AAA protocol (Chapter 2.1). Unmodified cysteine was not detectable by the method applied.

Figure 4.3 shows the determined single amino acid content for 8 g/L of AE (ISOGRO) and 8 g/L of *Pichia* YE in comparison to 20 % IPL-41 medium. The latter had been established in prior experiments (Chapter 2.1) as the amino acid reference which promotes cell viability and high protein yields in Sf9 insect cells. In line with the preparation of the extracts by acid hydrolysis, asparagine, glutamine and tryptophan are not present in the AE. Furthermore, the overall content of amino acids was comparable in both extracts, but differed for individual amino acids.

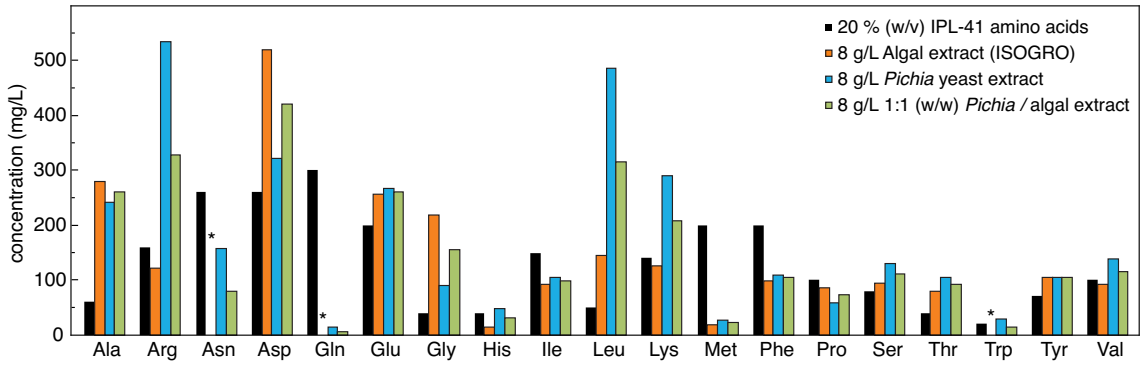


Figure 4.3 Amino acid content of different cell extracts and IPL-41. The free amino acid content of the commercial algal extract (orange) and the *Pichia* yeast extract (blue) was quantified by HPLC analysis following pre-column derivatization with phenyl isothiocyanate. Levels of cysteine were not determined by this method. The amino acid content of the both is compared to 20 % concentrations of amino acids in the IPL-41 medium (black), which was found to be sufficient to promote protein expression in Sf9 insect cells. In addition, the concentrations of amino acids yielded by blending both extracts (green) at 1:1 (w/w) ratio are given. The asterisk indicates missing amino acids in the algal extract.

Figure 4.3 also indicates that the content of aspartic acid is 70 % larger in AE than in YE. However, the combined content of aspartic acid and asparagine is similar in both extracts. This behaviour may be explained by the conversion of asparagine to aspartic acid in AE by the harsh acid hydrolysis conditions (*32*, *136*). Moreover, glutamine showed the lowest abundance of all amino acids in the YE and was not detectable in AE, which may relate to the instability of glutamine in solution. Depending on pH and temperature, glutamine is converted to glutamic and pyroglutamic acid to a different extent (*137*). Hydrolysis of algal cells by strong mineralic acids at low pH leads to the conversion of glutamine to glutamate (*135*). Hence, the amount of glutamate determined in the AE contains a fraction of glutamine initially present in the algal cells. In contrast, autolysis of *Pichia* was performed for five days at pH 6 and 50 °C. These milder conditions may primarily convert glutamine to pyroglutamic acid in the YE (*138*, *139*). However, pyroglutamic acid is not detectable by PTH-AAA without additional treatment of the sample (*140*). Furthermore, comparable levels of glutamate were observed for both extracts and thus the conversion of initial glutamine to glutamate in the final YE cannot be excluded (Chapter 2.1).

In addition, compared to AE significantly increased levels of arginine, histidine, leucine and lysine were observed for the YE, while the content of glycine was reduced. These differences may be attributed to the source organism and the method of preparation. In particular, it is expected that the differences in the used enzymatic and chemical hydrolysis protocols lead to variations in the amino acid yield.

The 1:1 mixture of YE and AE was sufficient to supply most of the amino acids in concentrations that exceed those of the 20 % IPL-41 medium (Figure 4.3). This observation is in line with the achieved expression yields (Figure 4.2) and explains why no further supplementation of tryptophan was required for the combined labeling scheme. As has been discussed in Chapter 2, the low abundance of methionine is apparently compensated by the metabolic capacity of the insect cell (*36*).

Solution state NMR spectroscopy of uniformly ^{15}N-labeled LBP

To proceed with NMR analysis, ^{15}N-labeled hLBP was expressed in Sf9 insect cells using a 1:1 ratio of ^{15}N-labeled AE/YE. The complete labeling procedure was adapted from the protocol that was originally developed for homemade YE involving an initial starvation step (Chapter 2.1). Thus, ΔSF4 supplemented with 1 g/L of unlabeled glutamine, 4 g/L of ^{15}N-labeled AE (N-ISOGRO, Sigma) and 4 g/L of ^{15}N-labeled YE was used as described before. Following purification, 2.7 mg/L of ^{15}N-labeled hLBP were obtained. Since it was previously observed that hLBP concentrations exceeding 8 mg/ml led to irreversible precipitation of the protein, only a fraction of 1.5 mg was used to prepare the NMR sample. Figure 4.4 shows the resulting ^{1}H-^{15}N TROSY-HSQC recorded for 60 h on the 111 µM ^{15}N-labeled hLBP. In total, about 324 of the 473 expected backbone amide resonances were observable. The determined amide proton T_2 value was 5.1 ms. With an estimated molecular weight of 52.8 kDa and additional glycosylation sites, further improvement of the spectral quality requires deuteration of the protein. Furthermore, ^{15}N-incorporation of hLBP was analyzed by a 1D spin echo difference experiment yielding an estimated ^{15}N-incorporation of 79 %.

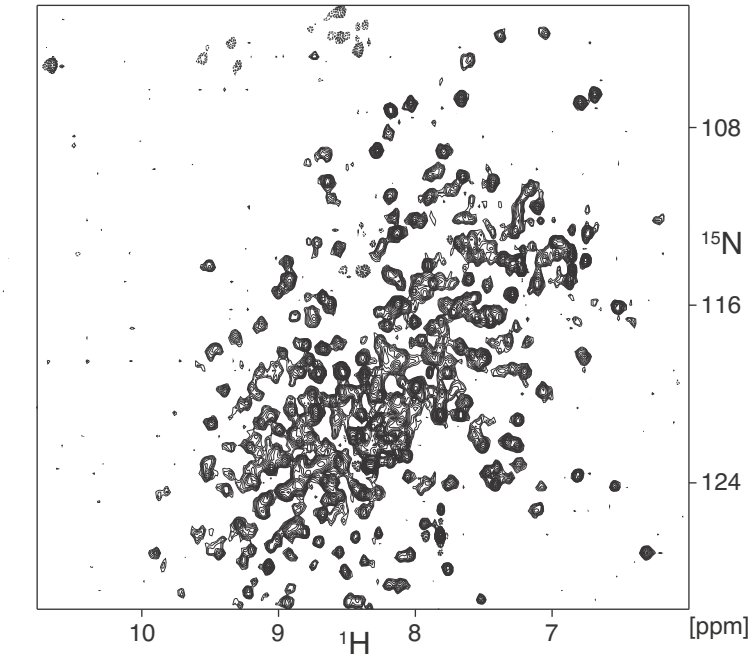


Figure 4.4 ^{1}H,^{15}N-TROSY-HSQC spectrum of ^{15}N-labeled human lipopolysaccharide-binding protein. ^{15}N-labeled hLBP was expressed in Sf9 cells grown on a combination of ^{15}N-yeast extract and ^{15}N-algal extract supplemented with unlabeled glutamine as described in the text. The TROSY-HSQC was recorded in 60 h on a 111 µM sample of ^{15}N-labeled hLBP.

So far mass spectrometric analysis on hLBP could not be established to quantify the incorporation of ^{15}N. However, a theoretical estimation of the labeling efficiency can be based on the previously obtained mass spectrometry data for cytosolic proteins (Chapter 2.1). When applied for expression of cytosolic proteins in Sf9 cells, the applied labeling scheme yielded an overall ^{15}N-incorporation of about 82 % for all non-glutamine residues as determined by mass spectrometric analysis. Therefore, the ^{15}N-incorporation of 79 % observed for hLBP by NMR is in good agreement with the mass spectrometric analysis of the cytosolic proteins cAbl and trGFPuv. Moreover, supplementation of ^{15}N-labeled YE and $^{15}N_2$-

glutamine may increase the average ^{15}N-incorporation of hLBP to 90 % as was observed for cytosolic proteins.

Conclusions and perspectives

The limitations of the developed yeastolate-based labeling protocol applied to a secreted protein have been investigated. Supplementation of homemade YE resulted in the lowest expression level of hLBP. However, robust expression yields were achieved in full SF-4 medium that contains commercial yeastolates. Unfortunately, the exact concentration in the medium and the preparation of those yeastolates is not disclosed.

In line with the recently reported labeling protocol (*39*), high yield expression of hLBP was obtained by supplementation of AE, tryptophan and glutamine into an amino acid depleted medium. Furthermore, the level of ^{15}N-incorporation achieved by combined application of AE and YE was similar to the AE-based protocol despite addition of unlabeled glutamine. This may be attributed to the additional starvation step used in the YE labeling protocol (Chapter 2.1). Therefore, it is expected that the reported incorporation levels of the AE-based approach may be improved by this starvation step.

Comparison of the algal and the yeast extract by amino acid analysis revealed only minor differences of the overall amino acid content. The differences for individual amino acids can be explained to a large extent by the method of lysate preparation. In particular, the loss of tryptophan and the conversion of glutamine and asparagine are direct results of acid hydrolysis used for AE. It is hoped that milder conditions for the preparation of algal cell lysates can be established to optimize the recovery of these amino acids. This may be achieved by straightforward adaptation of the autolysis protocol that was applied to *Pichia*.

Remarkably, the observed decrease in expression yield of hLBP in YE based labeling medium is not caused by significant differences in the amino acid composition and the related metabolism of amino acids in the insect cell. Efficient secretion of hLBP may be influenced by several factors related to the YE composition. In particular, YE and AE differ significantly in their respective content of mono- and polysaccharides. Commercial protocols for the production of AE involve removal of nearly all carbohydrates prior to acid hydrolysis of the residual biomass. As compared to AE, autolysates of yeast contain carbohydrates, which may link the observed behavior with *N*-glycosylation of the precursor protein. *N*-glycosylation was identified as a strict requirement to enter the secretory pathway (*141*). However, refined analysis of both extracts to characterize carbohydrates and components other than amino acids is required to understand the difference in expression yield of AE compared to YE (*142*, *143*).

Moreover, it has been shown that uniform ^{15}N-labeling of hLBP in insect cells is possible by a combination of ^{15}N-labeled cell lysates of algae and yeast providing provides a basis of an economic labeling medium for expression of secreted proteins in insect cells. This initial labeling scheme can be expanded to uniform ^{13}C and fractional ^{2}H labeling in a straightforward manner. Deuteration has the prospect to further increase the spectral quality and thus will enable a more detailed characterization of hLBP and its interaction with lipopolysaccharide (LPS) by solution and solid state NMR experiments.

In a similar approach, bacterial and algal extracts had been successfully combined for the production of labeled proteins in mammalian cell lines (*32*). Furthermore, studies in Chinese hamster ovarian cells have already shown that supplementation of YE tremendously increases

cell proliferation and protein expression (*144*). The combined use of AE and YE has the potential to become an economic labeling strategy for mammalian cells.

Material and Methods

Isotope labeling and preparation of labeling media

All custom media and stable isotope labeling schemes based on cellular extracts of yeast and algae were prepared as described in Chapter 2.1. ^{15}N-labeled algal cell extract was purchased from a commercial supplier (ISOGRO, Sigma-Aldrich).

Cell culture and preparation of recombinant Baculovirus

In general, handling and maintenance of Sf9 cell culture was based on established protocols (*118*). Permanent cultures were subcultured in mid-log phase and kept in full SF-4 medium (Bioconcept) and Insect Xpress medium (Lonza) at 27 °C and shaken at 80 rpm. Recombinant Bacmid DNA was generated in DH10Bac *E. coli* cells using the commercial pFastBac vector according to the supplier's procedure (Life Technologies). Baculovirus was generated in adherent cultures as reported by O'Reilly et al. (*118*). Subsequently, high-titer virus stocks for expression were produced by two additional amplification rounds in suspension cultures. In the first amplification round, 50 ml of Sf9 cell culture containing $1x10^6$ cells/ml in serum-free medium were infected with recombinant virus and cultured until cell viability decreased to ≤90 %. The supernatant (P1) was collected by centrifugation at 1000*g* and 4 °C for 15 min. For the second amplification round, 100 ml of serum-free Sf9 cell culture containing $1x10^6$ cells/ml was infected with P1 and processed accordingly to yield a P2 virus stock.

Functional titration of the P2 virus stock was applied to determine the optimal amount for expression. Briefly, small-scale cultures of Sf9 cells were infected in full SF-4 medium at a cell density of $1.5x10^6$ cells/ml with P2 virus stock corresponding to functional titers of 1 to 10 ml per 1 L of cell culture. The yield of purified protein was then quantified for different expression times.

Cloning and purification of human lipopolysaccharide binding protein

Leonildo Delgado is gratefully acknowledged for establishing the expression and purification protocol of human lipopolysaccharide binding protein (hLBP). A synthesized gene sequence of hLBP was purchased from a commercial supplier (Genscript). Recombinant Baculovirus was prepared by insertion of the hLBP gene into the pFastBac vector (Life Technologies). Bacmid DNA was produced in DH10Bac *E. coli* cells according to the standard protocol provided by the supplier. Functional virus titer and the optimal expression time were determined in small-scale expression tests. hLBP was subsequently expressed in SF4-adapted Sf9 insect cells until the cell viability decreased to 80-85 %.

After expression, medium containing secreted hLBP was seperated from the cells by centrifugation at 1000*g* for 10 min (4°C). The supernatant was filtered through a 0.22 µm filter and directly loaded on a strong cation exchange column (SP-Sepharose Fast Flow 2 × 5 ml, GE Healthcare Life Sciences) equilibrated with 25 mM HEPES (pH 7.5), 100 mM NaCl and 5% glycerol (buffer A). Following a washing step with 2 column volumes (CV) of Buffer A, hLBP was eluted using a gradient of 25 mM HEPES (pH 7.5), 1 M NaCl and 5 %

glycerol (buffer B) from 0-100 % over 20 CV. Fractions containing hLBP were combined and subsequently loaded onto a Ni Sepharose column (HisTrap HP 5 ml, GE Healthcare Life Sciences) initially equilibrated with 25 mM HEPES (pH 7.5), 500 mM NaCl and 5% glycerol (buffer A). The column was washed with 2 CV of buffer A, followed by gradient from 0-5 % in 4 CV of 25 mM HEPES (pH 7.5), 500 mM NaCl, 500 mM imidazole and 5 % glycerol (buffer B). Bound hLBP was finally eluted by a gradient of buffer B from 5-100 % in a total volume of 4 CV. Fractions containing hLBP were combined and concentrated by centrifugation in a 30 kDa MWCO filter (VivaSpin 15, Sartorius). The concentrated supernatant was loaded onto a size-exclusion column (Superdex 200 10/300 GL, GE Healthcare Life Sciences) equilibrated with 5 mM HEPES (pH 7.5), 500 mM NaCl and 1 mM EDTA. Fractions containing hLBP were combined and concentrated for further analysis (VivaSpin 15, Sartorius).

The NMR sample of ^{15}N-hLBP was prepared in 5 mM HEPES (pH 7.4), 500 mM NaCl, 1 mM EDTA, 5 % D_2O and 0.01 % NaN_3.

NMR spectroscopy and data processing

NMR experiments were recorded on a Bruker Avance III 900 MHz spectrometer equipped with a triple-resonance cryoprobe. All experiments on hLBP were acquired at a temperature of 37 °C. Amide proton T_2 values were determined from a jump-return spin-echo experiment (*145*). The 2D ^{1}H-^{15}N-TROSY experiment on ^{15}N-hLBP was recorded with 130 (^{15}N) x 1024 (^{1}H) complex points, acquisition times of 26 ms (^{15}N) x 43 (^{1}H) ms, and a total experimental time of 60 h. INEPT transfer times were optimized for amide proton T_2 values of hLBP. NMR data were processed and analyzed with the NMRPipe software package (*128*).

Yeast growth and preparation of yeast extract

Detailed protocols are presented in Chapter 2.1.

Quantitative amino acid analysis by LC/MS

Detailed protocols are presented in Chapter 2.1.

5 Synthesis of isotope-labeled glutamine by a two-step enzymatic procedure

Introduction

L-glutamine plays an essential role for labeling media based on yeast and algal cell extracts described in Chapter 2.1 and Chapter 3. Insect cells utilize L-glutamine within the tricarboxylic acid (TCA) cycle, where it functions as an important nitrogen source for the synthesis of non-essential amino acids (*35*, *146*, *147*). In particular, L-glutamine is used to overcome limitations in L-glutamate, L-aspartate and L-asparagine (*30*). Consequently, this metabolic flexibility leads to significant isotope scrambling for ^{15}N-labeled L-glutamine. While addition of L-glutamine to the medium is not strictly required for the growth of insect cells (*35*, *37*), supplementation of isotope-labeled L-glutamine ensures sufficient expression levels and enables high isotope incorporation in insect cell culture media (Chapter 2.1). Unfortunately, all labeled variants of L-glutamine are expensive due to the instability of the amino acid. As indicated in Chapter 2.1, the costs of isotope labeling by ^{15}N or ^{2}H,^{15}N are dominated by the price of $^{15}N_2$-L-glutamine, which amounts to about 82 % of the total costs of the yeastolate-based labeling medium.

Over the last decades, the use of deuteration has played a key role in advanced biomolecular NMR studies. Indeed, the majority of backbone and side-chain NMR experiments greatly benefit from deuteration especially when applied to larger proteins (*148*, *149*). However, the incorporation of deuterons into proteins produced in the baculovirus/insect cell expression system is difficult. The present method (Chapter 2.1) of supplementing homemade deuterated yeastolates and protonated L-glutamine in the labeling medium has only led to 60-70 % deuteration. Although the overall level of deuteration is expected to improve by supplementation of deuterated L-glutamine, it comes at a prohibitive cost that might prevent routine application for many laboratories.

As an alternative to purchasing isotope-labeled L-glutamine, it is also possible to synthesize L-glutamine from 2-oxoglutaric acid by enzymatic reactions in a laboratory environment. The first step involves reductive amination of 2-oxoglutaric acid by glutamate dehydrogenase to L-glutamic acid, which is then used as the precursor for the synthesis of L-glutamine. A repetitive batch setup for the synthesis of L-glutamic acid from 2-oxoglutaric acid was reported two decades ago (*150*). The second step, which involves the enzymatic conversion of L-glutamic acid to L-glutamine upon condensation with ammonium, is catalyzed by a class of enzymes known as glutamate-ammonia ligases (*151*). This reaction is widely established for biochemical assays, but has seen only limited use for preparative scale production of L-glutamine so far (*32*).

In this chapter, we present a detailed protocol for the enzymatic synthesis of L-glutamine which enables isotope labeling by ^{15}N including fractional deuteration in an economic manner. Figure 5.1 depicts the overall reaction scheme, which combines the repetitive batch synthesis of L-glutamic acid (1) with the subsequent enzymatic conversion to L-glutamine (2). Fractional deuteration of L-glutamic acid was achieved by deuteration of the precursor 2-oxoglutaric acid and analyzed by NMR spectroscopy. Furthermore, recombinant expression and purification of glutamate-ammonia ligase has been carried out to obtain a functional form of the enzyme thereby reducing the overall costs.

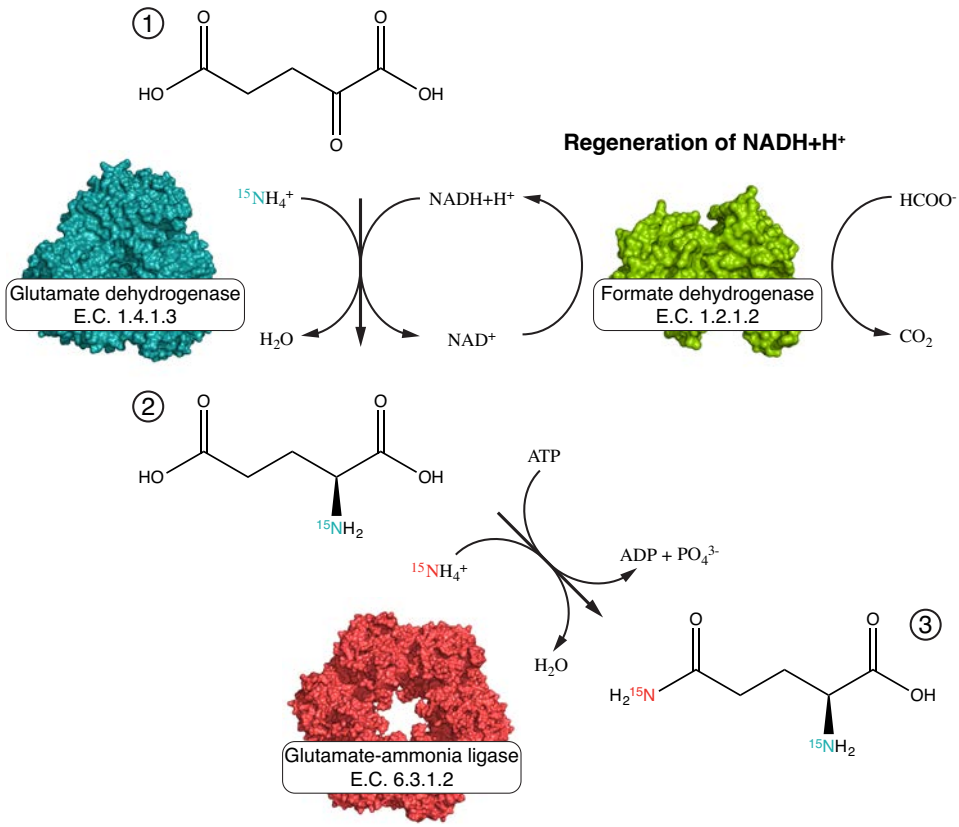


Figure 5.1 Enzymatic synthesis of isotope-labeled L-glutamine. In a first step, 2-oxoglutaric acid (1) and NH_4Cl are converted to L-glutamic acid (2) by glutamate dehydrogenase. This reaction is coupled to the regeneration of $NADH+H^+$ by formate dehydrogenase. Subsequently, (2) and NH_4Cl are converted to $^{15}N_2$-L-glutamine (3) by glutamate-ammonia ligase. In both reactions, use of $^{15}NH_4Cl$ leads to incorporation of ^{15}N at the amine and amide positions.

Results and Discussion

Synthesis of ^{15}N-L-glutamic acid from 2-oxoglutaric acid and NH_4Cl in a repetitive batch setup

L-glutamic acid was produced by condensation of 2-oxoglutaric acid with NH_4^+ by glutamate dehydrogenase (GDH) according to the scheme depicted in Figure 5.1. As indicated in this scheme, efficient regeneration of NADH was achieved through oxidation of formate by formate dehydrogenase (FDH), which was reported previously by Kragl and coworkers (*150*). A detailed protocol of the reaction is outlined in Table 5.1.

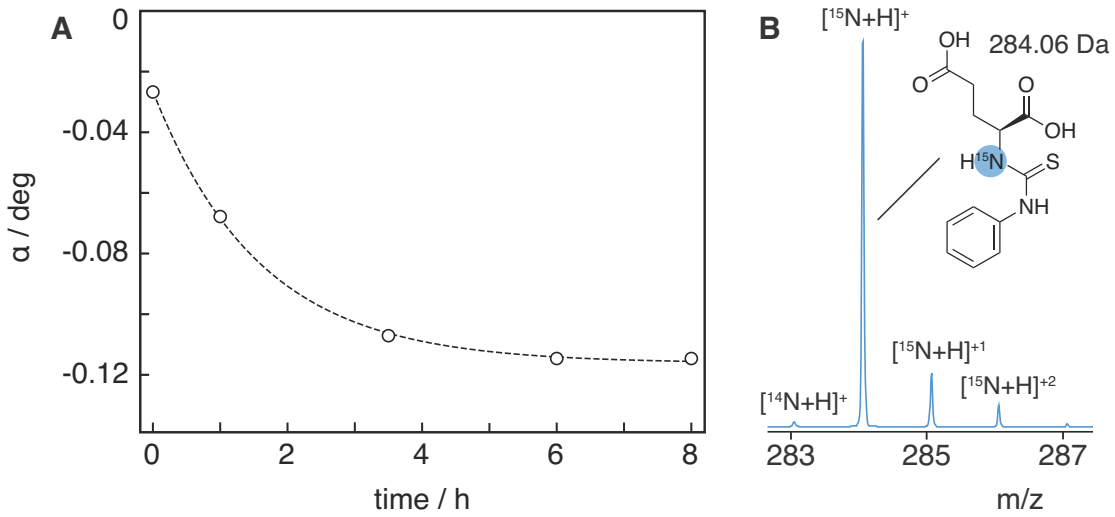


Figure 5.2 A Polarimetric analysis of L-glutamic acid formation at 589 nm. ORD was directly measured in the enzymatic reaction (Tab 5.1) at 25 °C using a light path of 1 dm yielding a specific rotation $[\alpha]_D^{25}$ of -1.24 deg·dm^{-1}·ml·g^{-1}. The dashed line indicates the obtained fit for a single exponential decay function. Degrees of rotation (α) were related to the concentration of L-glutamic acid by a calibration curve. **B** Incorporation of ^{15}N determined by ESI-TOF mass spectrometry. L-glutamic acid was derivatized with phenyl isothiocyanat (PTC) prior to analysis by LC/MS leading to a monoisotopic mass of 283.06 Da. Likewise a mass of 284.06 Da corresponds to the ^{15}N-labeled PTC-L-glutamic acid.

L-glutamic acid is a chiral molecule and displays optical activity. The ability to rotate the plane of linearly polarized light is known as optical rotation and in the case of L-glutamic acid is strongly governed by the allowed electronic transitions of the carboxylate chromophore. Optical rotatory dispersion (ORD) can be measured at wavelengths far removed from the resonance frequency of the carboxylate group due to the dispersive nature of the phenomenon. To this purpose, samples were taken during course of the reaction and optical rotation was determined at 589 nm (Sodium D line), 25 °C and pH 8. A calibration curve for L-glutamic acid standards covering the concentration range of the reaction was determined by linear regression.

Figure 5.2A shows the obtained degrees of rotation (α) over the course of the reaction. The initial value at the time point t = 0 h was determined in the absence of GDH and FDH, indicating the optical activity of NAD^+ in the reaction mix. However, this contribution to the overall optical activity of the reaction mixture may change over the time course, as accumulation of NADH by activity of FDH occurs at the end of the reaction. To determine

the rate of product formation, a rapid equilibrium mechanism was assumed (*152*) and the data was fitted to a single exponential decay yielding an apparent rate constant of 0.63 h^{-1} and an amplitude of 0.09 degrees (dashed line, Figure. 5.2A). Using concentrations of 20 U/ml GDH and 200 mM 2-oxoglutaric acid (Table 5.1), this corresponds to a catalytic rate of 0.26 mM^{-1} U^{-1} h^{-1}. This rate is in good agreement with the expected value of 0.28 mM^{-1} U^{-1} h^{-1} based on the unit definition indicated by the commercial supplier (one unit converts 1.0 µmole of 2-oxoglutaric acid to L-glutamic acid per min at pH 7.3, 25 °C).

Furthermore, the time course of the reaction indicates that the equilibrium was reached after 6 h to 8 h. However, the determined α values correspond to a final yield of only 88.6 %. In a similar approach, the yield of L-glutamic acid was reported to be >95 % within 4 h of reaction (*150*). Both, the slight decrease in the rate of product formation and the yield suggest that the reaction conditions were not yet optimal. Interestingly, substrate inhibition of GDH by NADH was reported for a concentration ratio of the enzyme and the cofactor at similar conditions (*152*). This may relate to the activity of FDH, which determines the ratio of oxidized and reduced NAD(H). Therefore, increasing the initial amount of GDH or decreasing the initial amount of FDH may increase the rate and the yield of the reaction. Alternatively, simple reduction of the initial formate concentration may improve the reaction kinetics.

Table 5.1 Repetetive batch synthesis of ^{15}N-L-glutamic acid from 2-oxoglutaric acid and $^{15}NH_4Cl$

Enzymatic synthesis of $^{15}N_2$-L-glutamic acid	
Solution 1	1. Prepare 100 ml of solution 1 in ultrapure water (Milli-Q) as follows: - 2.92 g 2-oxoglutaric acid (200 mM) - 1.09 g $^{15}NH_4Cl$ (200 mM) - 2.72 g sodium formate (400 mM) 2. Adjust solution to pH 8 3. Filter-sterilize and store at 4 °C
Enzymatic reaction	1. Prepare reaction mix in 20 ml solution 1 as follows: - 1 mM NAD - 20 U/ml glutamate dehydrogenase (GDH) - 1 U/ml formate dehydrogenase (FDH) 2. Run reaction at 25 °C in a stirred filtration cell equipped with a 10 kDa MWCO filter membrane (e.g. AMICON) 3. Follow reaction by polarimetric analysis at 589 nm 4. Upon completion, filter reaction mix through the 10 kDa MWCO and collect permeate for crystallization 5. Keep retentate (typically 1-2 ml) and add 18-19 ml of solution 1 to continue the reaction (step 2)
Crystallization	1. Combine permeates from all repetitive runs 2. Adjust pH of combined permeate to 3.2 3. Cool solution to 4 °C and let crystalize overnight 4. Separate crystals from residual permeate by filtration (glass frit, fine porosity) 5. Wash crystals with ice-cold Milli-Q water 6. Dry crystals in desiccator 7. Perform another crystallization round with the residual permeate (step 4) using the obtained L-glutamic acid crystals as seeds 8. Analyze purity and isotope incorporation by LC/MS

To obtain pure L-glutamic acid, the reaction mix was filtered and the obtained permeate was subjected to crystallization as described in Table 5.1. In a first step, 76 % of the theoretical yield was obtained. An additional 12 % of crystalline L-glutamic acid was

recovered in subsequent crystallization step, amounting for a total yield of 88 %. This is in good agreement with the value determined by polarimetric analysis.

Incorporation of ^{15}N into synthesized L-glutamic acid was analyzed by mass spectrometry (Figure 5.1B), revealing a residual unlabeled fraction of 1.02 %, which was calculated as the ratio of the unlabeled peak integral over the sum of all observed ^{15}N-labeled species. This value is consistent with the ^{15}N-incorporation of $^{15}NH_4Cl$ (99 %). Likewise, mass spectrometry analysis showed that no traces of 2-oxoglutaric acid were present following crystallization.

Enzymatic conversion of L-glutamate to L-glutamine by glutamate-ammonia ligase

In the second step of the enzymatic synthesis, ^{15}N-L-glutamic acid was converted to $^{15}N_2$-L-glutamine by a commercial glutamate-ammonia ligase (GAL) isolated from *E. coli*. A detailed protocol is presented in Table 5.2. Again, $^{15}NH_4Cl$ was employed as the nitrogen source in the reaction mixture. Thus, the condensation of $^{15}NH_4^+$ with L-glutamic acid will incorporate ^{15}N into the carboxamide group of L-glutamine. The enzymatic reaction was analyzed by thin-layer chromatography (TLC) developed in 20:5:1 isopropanol/water/formic acid. Commonly, a silica gel matrix is used to separate L-glutamine and L-glutamic acid based on their different adsorption behavior. This difference in adsorption is caused by the distinct polarity of the functional groups in the side chains of both amino acids. Compared to the carboxamide group of L-glutamine, the more polar carboxylic group is decreasing the hydrophobicity of L-glutamic acid (*153*). Consequently, L-glutamine will interact stronger with the silica gel matrix of the TLC plate. A quantitative measure of the interaction and hence a way to identify different compounds is given by the retention factor, which is equal to the distance migrated by the compound over the distance covered by the solvent following complete development.

Table 5.2 Conversion of ^{15}N-glutamic acid and to $^{15}N_2$-L-glutamine by glutamate-ammonia ligase

Production of $^{15}N_2$-L-glutamine by glutamate-ammonia ligase[a]
1. Prepare 100 ml of reaction mix in ultrapure water (Milli-Q) as follows: - 1 g ^{15}N-L-glutamic acid (67 mM) - 0.36 g $^{15}NH_4Cl$ (67 mM) - 6.6 g ATP-2Na (120 mM) - 1.9 g $MgCl_2 \times 6\ H_2O$ (93 mM), *for optimal reaction speed 120 mM should be used* - 6 U/ml of glutamate-ammonia ligase (GAL), *12 U/ml were required to complete the reaction* (Figure 5.3) - 50 µg/ml gentamycin
2. Adjust to pH 7.6 and filter-sterilize.
3. Run reaction at 37 °C for 48 h to 72 h.
4. Maintain pH at 7.6 by addition of NaOH.
5. Analyze reaction by thin-layer-chromatography or LC/MS.

[a] Optimized reaction conditions are described in the main text

Figure 5.3 shows the developed TLC plates for different reaction conditions including a reference for each amino acid. Retention factors of 0.54 and 0.42 were determined for commercial standards of L-glutamic acid and L-glutamine, respectively (Lane 1 and 2, Figure 5.3). Initially, the enzymatic reaction was carried out in batch-mode at pH 7.5 and 37 °C for 4 days without readjustment of the pH. Analysis of the intensity ratio obtained by TLC revealed that only a fraction of about 40 % was converted to L-glutamine under these

conditions (Lane 3, Figure 5.3). Further extension of the reaction time to a total length of 7 days did not increase the yield of L-glutamine. However, a decrease in pH was observed during the course of the reaction. In the initial setup, the pH of the reaction mixture decreased to 6.5 after 16 h. The pH dependence of the enzyme activity was well described in previous studies, showing an optimum at pH 7.5 (*154*). With decreasing pH the activity significantly decreases, having declined to less than 60 % at pH 6.5 (*154*).

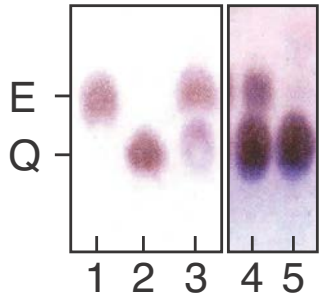


Figure 5.3 Enzymatic synthesis of L-glutamine analyzed by thin-layer chromatography. Commercial standards of E (1) and Q (2) were employed to establish retention factors. Initial reaction conditions without readjustment of the pH (3) and optimized reaction under continuous pH control (4). Complete conversion to L-glutamine upon two-fold increase in enzyme concentration and tight pH control (5).

To counteract this decrease and the correlated loss of enzyme activity, the pH of the reaction mixture was monitored and readjusted to 7.5 by addition of 5 M NaOH at regular time points during the course of the reaction. Without further modification of the reaction mixture, it was possible to increase the yield of L-glutamine to about 70 % after 2 days of reaction according to the intensity ratio of the TLC spots (Lane 4, Figure 5.3). The reaction was thus continued for another 2 days amounting to a total reaction time of 4 days. Remarkably, the yield remained unchanged indicating another limiting factor other than pH optimum. Therefore, the amount of GAL in the reaction mixture was varied to assess the effect of the enzyme concentration. Upon increasing the concentration of GAL by two-fold, complete conversion of L-glutamic acid to L-glutamine was achieved within 2 days (Lane 5, Figure 5.3). This finding may suggest a relation of the incomplete conversion to product inhibition by accumulation of ADP and pyrophosphate. In addition, the initial concentration of ATP has a significant impact on the velocity of catalyzed reaction (*146*). When free ATP is present in concentrations exceeding 30 mM, a decrease in the reaction rate was observed. The reaction was carried out using 93 mM of $MgCl_2$ (Table 5.1), whereas an equimolar concentration of divalent magnesium ions is required to prevent substrate inactivation caused by free ATP. Therefore, reaction speed and yield may increase by addition of 120 mM $MgCl_2$ to account for the concentration of ATP.

Since the entire amount of L-glutamic acid was converted to L-glutamine in this step of the reaction, the overall yield of L-glutamine production is simply limited by the efficiency of the first reaction, where 88 % of 2-oxoglutaric acid was converted to L-glutamic acid by GDH. Hence, the combined efficiency of the reaction is equal to 88 %.

Fractional deuteration of L-glutamic acid by proton-deuterium exchange

The described enzymatic synthesis of L-glutamine based on the precursor 2-oxoglutaric acid is amenable to partial deuteration. A detailed protocol for the selective deuteration of L-glutamic acid at the H^{β} positions is given in Table 5.3. In brief, the precursor 2-oxoglutaric acid was incubated at pD 10, 45 °C in 99.8 % D_2O for 16 h to form 2-oxoglutaric-3,3'-d_2 acid. A similar approach was applied to achieve deuteration of 2-oxobutyric acid for the generation of ^{13}C,^{2}H-(^{1}H-δ1 methyl)-isoleucine for NMR studies (*2*).

Table 5.3 Fractional deuteration of 2-oxoglutaric acid in D_2O

Preparation of 2-oxoglutaric-2,2'-d_2 acid
1. Prepare 10 ml of reaction mix (RM) containing 1 g of 2-oxoglutaric acid in 99.8 % D_2O
2. Adjust RM to pD 10 by addition of NaOD in D_2O
3. Incubate for 16 h at 45 °C
4. Lyophilize RM
5. Add lyophilized powder to Solution 1 (see Table 5.1)

The underlying mechanism based on the keto-enol tautomerism of 2-oxoglutaric acid is depicted in Figure 5.4. As shown in this scheme, the acidity of the protons at the C3 position allows for the exchange of these protons with the deuterated solvent. In the next step, enzymatic conversion of 2-oxoglutaric-3,3'-d_2 acid by activity of GDH results in selective deuteration of the H^{β} positions in L-glutamic acid, which enables production of selectively deuterated L-glutamine according to the protocol given in Table 5.2.

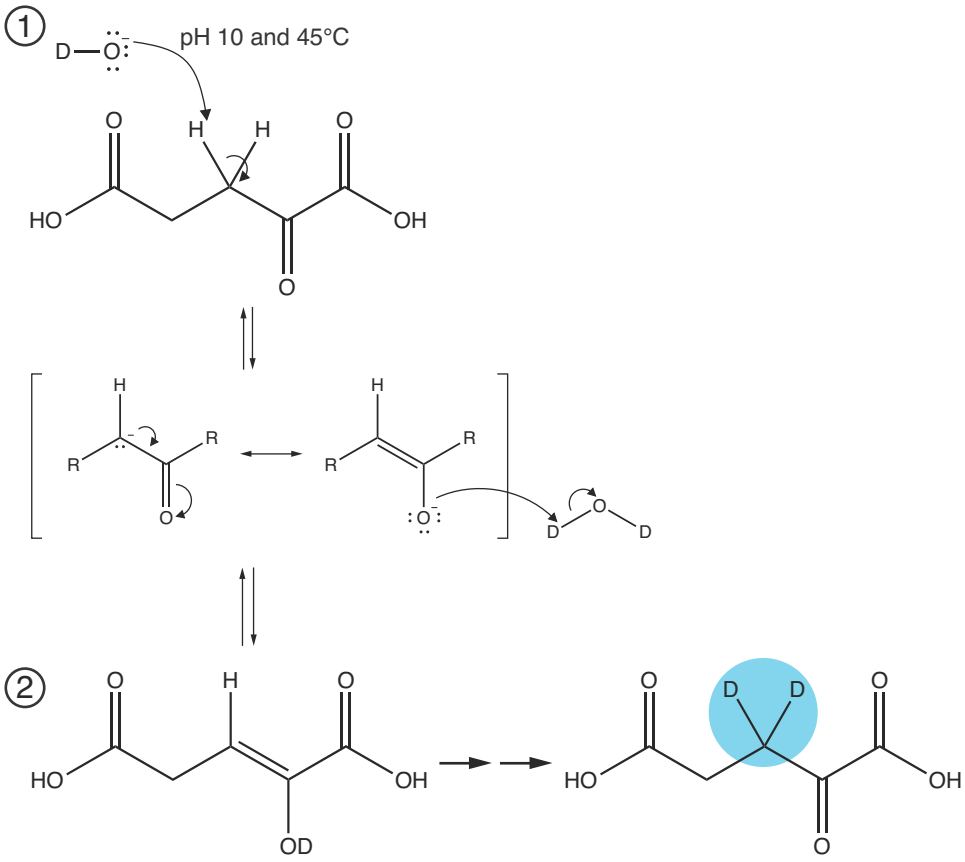


Figure 5.4 Base-catalyzed proton-deuterium exchange reaction of 2-oxolgutaric acid. The depicted keto-enol tautomerism (1) at elevated temperature and pH imposes an equilibrium of both tautomeric forms leading to an enolization of 2-oxoglutaric acid. Hence, the initial protonation of the C3 position is lost. The backward reaction when carried out in D_2O eventually enables deuteration of the C3 position (blue circle) to yield 2-oxoglutaric-3,3'-d_2 acid (2).

Analysis of deuterium incorporation into L-glutamic acid by NMR

Incorporation of deuterium at different positions of L-glutamic acid was analyzed by 1D ^{1}H NMR spectroscopy based on comparison with the fully protonated amino acid (Figure 5.5). A reference spectrum was obtained for protonated L-glutamic acid that displays splitting of the proton lines due to homonuclear J_{HH}-couplings (Figure 5.5A). The assignment of the H^{α} (orange), H^{β} (green) and H^{γ} protons (blue) was derived from the ^{1}H chemical shift and expected multiplet structure (Figure 5.5B). In the case of L-glutamic-3,3'-d_2 acid where homemade 2-oxoglutaric-3,3'-d_2 acid prepared according to the exchange protocol in Table 5.3 was employed as a precursor, the resonances of the H^{β} protons disappear from the spectrum (Figure 5.5C). In addition, the H^{α} multiplet collapsed into a single line indicating efficient deuteration of the H^{β} position. This is also evident from the obtained doublet of doublets of the H^{γ} protons, which originates from the $^{1}J_{HH}$ to the second H^{γ} proton and the $^{4}J_{HH}$ coupling to the H^{α} proton, respectively. Furthermore a secondary isotope shift of about 12 ppb was observed for the H^{α} proton upon deuteration of the H^{β} position.

To quantify the level of deuteration in the selectively deuterated samples, the residual H^{β} and H^{γ} resonances were integrated and the resulting integrals calibrated on the TSP signal. Proton-deuterium exchange (PDE) as described in Table 5.3 results in deuteration of the H^{β} positions, whereas the H^{γ} positions remain protonated. Following PDE, a residual protonation level of about 3.9 % was observed for the H^{β} position. This is slightly higher than the estimated residual protonation level in the reaction mixture of about 1.2 % based on the exchangeable protons of 2-oxoglutaric acid and may be explained by incomplete exchange.

Enzymatic conversion of the PDE-based precursor results in L-glutamic-3,3'-d_2 acid but leaves H^{α} and H^{γ} positions protonated. In an effort to further extend the level of deuteration, the commercial precursor 2-oxoglutaric-3,3',4,4'-d_4 acid was employed for enzymatic synthesis of L-glutamic acid according to Table 5.1. Conversion of this precursor will generate L-glutamic-3,3',4,4'-d_4 acid leading to the deuteration of the H^{β} and H^{γ} positions. The additional deuteration of the latter is shown 1D ^{1}H NMR spectrum of L-glutamic-3,3',4,4'-d_4 acid (Figure 5.5D). Obviously, the H^{γ} proton resonances disappear as a consequence of efficient deuteration. Again, the splitting of H^{α} proton line was absent due to substitution of the adjacent H^{β} protons by deuterons. The residual protonation levels of the H^{β} and H^{γ} positions were found to be 0.8 % and 2.1 %, respectively. These correspond to deuteration levels of 99.2 % and 97.9 %, which are in good agreement with the average deuteration level of 98 % of 2-oxoglutaric-3,3',4,4'-d_4 indicated by the commercial supplier. However, the H^{α} positions still remain protonated for both precursors, the commercial 2-oxoglutaric-3,3',4,4'-d_4 acid and the homemade 2-oxoglutaric-3,3'-d_2.

It is well understood that the relaxation of H^{N} and $^{13}C^{\alpha}$ is dominated by dipole-dipole interactions with the H^{α} proton (*126*, *155*). Hence, deuteration of the H^{α} position is expected to yield a significant increase in sensitivity for these nuclei in modern backbone and side chain NMR experiments. To replace the remaining H^{α} protons in L-glutamic acid, addition of formic-*d* acid would be required in the initial step of the described protocol. It is expected that this leads to the deuteration of NAD^{+} by FDH to form $NAD^{2}H$ (*156*). $NAD^{2}H$ would subsequently be used by GDH during reductive amination of 2-oxoglutaric acid allowing incorporation of deuterium at the H^{α} position of L-glutamic acid. Hence, the synthesis of uniformly deuterated L-glutamic-2,3,3',4,4'-d_5 as well as fractional deuterated L-glutamic-2,3,3'-d_3 acid may be established by this procedure using commercial

2-oxoglutaric-3,3',4,4'-d_4 acid and homemade 2-oxoglutaric-3,3'-d_2 acid, respectively. This approach has not yet been tested during the current thesis.

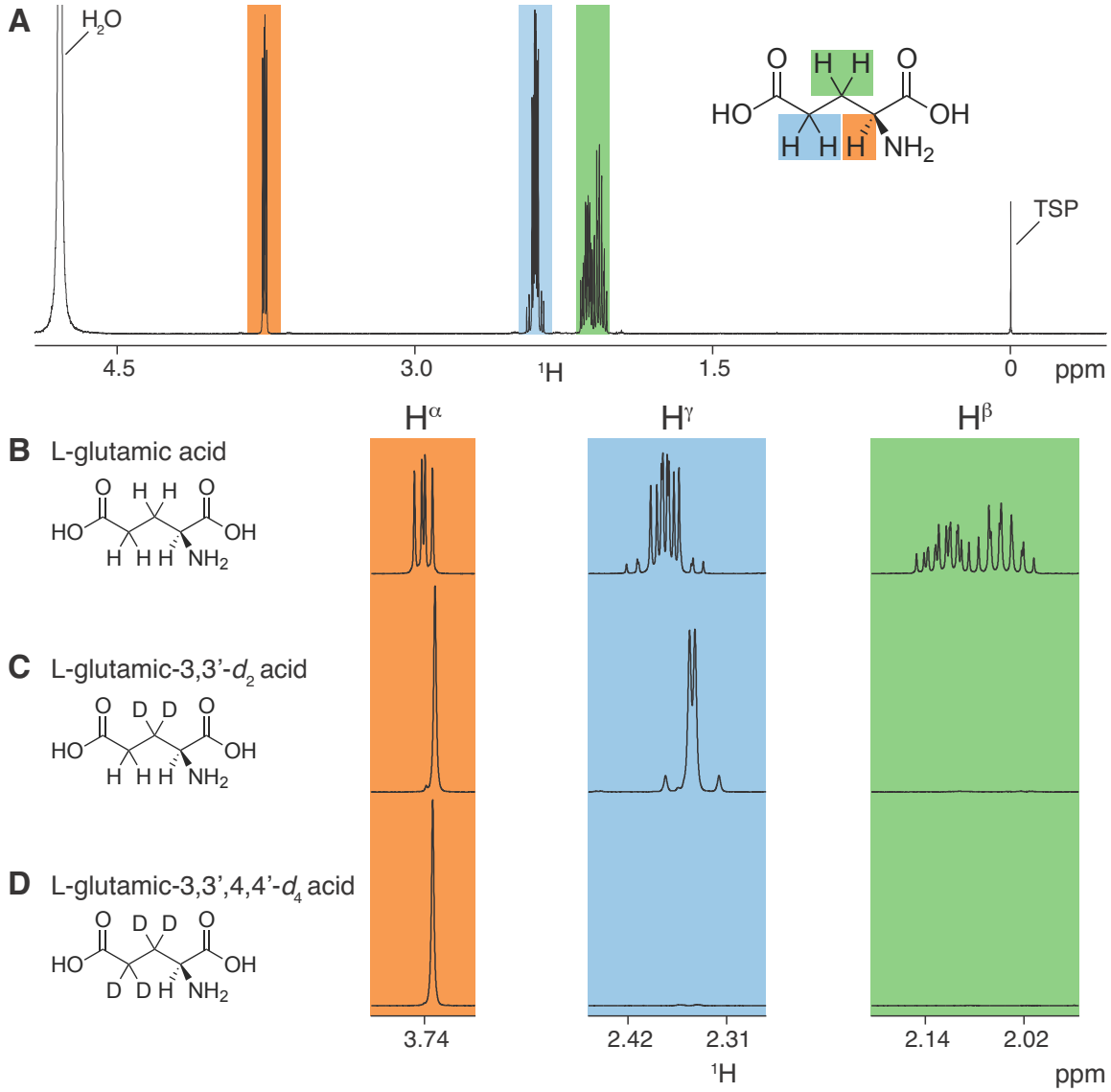


Figure 5.5 Deuteration of L-glutamic acid analyzed by 1D ^{1}H NMR spectroscopy. **A** The protonated reference indicating the H^{α} (orange), H^{β} (green) and H^{γ} (blue) resonances. *Bottom*: Detailed comparison of protonated (**B**), fractional (**C**) and perdeuterated L-glutamic acid (**D**). Preparation of fractional and perdeuterated L-glutamic acid was performed as described in the main text.

It is expected that deuteration of L-glutamic acid for all the described precursors of 2-oxoglutaric acid may be extended to L-glutamine, when deuterated L-glutamic acid is used for the synthesis of L-glutamine by activity of glutamate-ammonia ligase according to Table 5.2.

Recombinant expression and purification of glutamate-ammonia ligase

To further reduce the cost of the procedure, recombinant glutamate-ammonia ligase (GAL) of *Mycobacterium tuberculosis* was expressed in *E. coli*. Initial expression trials indicated that a fraction of GAL was present in the pellet after cell lysis. Therefore, GAL was expressed at 30 °C to reduce aggregation and increase the recovery of soluble protein. The soluble form of the protein was obtained by a two-step purification based on Ni^{2+}-NTA affinity and gel-filtration chromatography (see Materials and Methods) yielding 30 mg/L of purified GAL, which amounts to about 70 % of the total expression level when compared to remaining fraction in the pellet by SDS-PAGE analysis.

GAL, like other bacterial glutamate-ammonia ligases, assembles into a dodecamer in solution, which is formed by two hexameric rings, stacked on top of each other (Figure 5.6A). Here, each active site includes contributions from two adjacent subunits. This dodecameric quaternary structure is required for the activity of the enzyme. Multi-angle light scattering coupled to a gel-filtration chromatography (SEC-MALS) was employed to determine the mass of GAL at different concentrations. Figure 5.6B shows the obtained chromatograms and associated molar mass values. The determined mass of 665.54 ± 3.1 kDa indicates a native dodecameric state that was found to be stable within a concentration range of 0.25 mg/ml to 1.5 mg/ml. However, the mass of the dodecamer based on the recombinant sequence of GAL is expected to be 656.7 kDa. While this observed difference may be explained by inaccuracy of the calibration used for the SEC-MALS setup, it could as well be related to the described adenylylation (AMPylation) of GAL in *E. coli* (*157*).

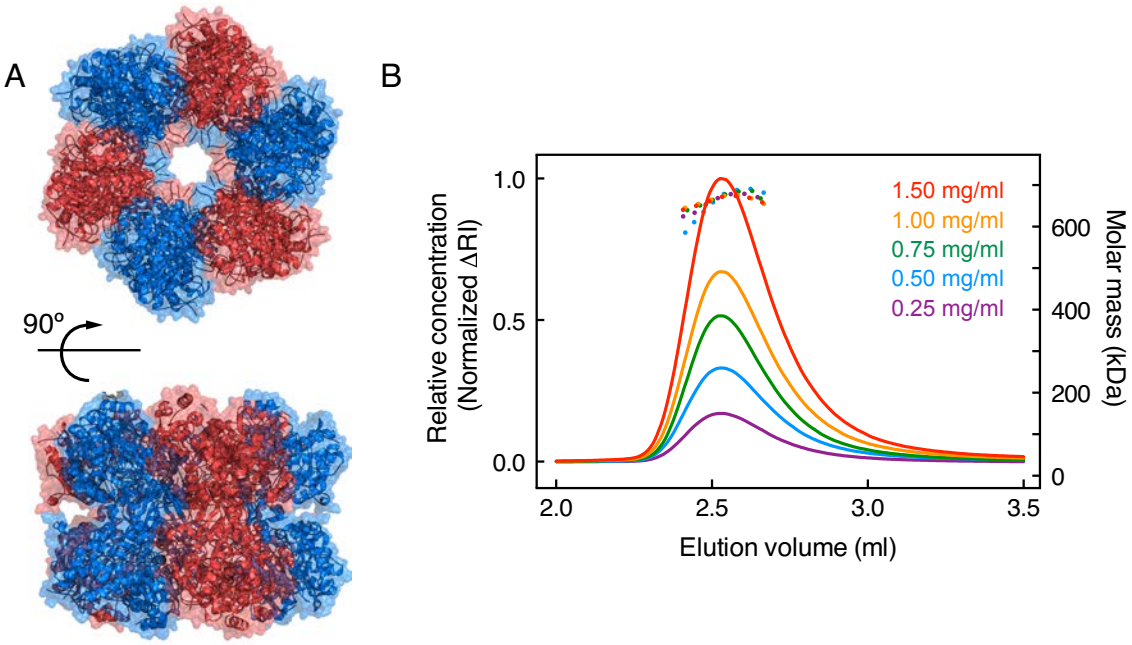


Figure 5.6 A Functional assembly of bacterial glutamine-ammonia ligase (GAL). The dodecameric quartenary structure is formed by two hexamers stacked on top of each other, consisting of six identical monomeric subunits. Adjacent subunits are alternatingly colored in red and blue. **B** Multi-angle light scattering chromatrograms of recombinat GAL from *Mycobacterium tuberculosis* expressed in *E. coli*. Multiple runs were performed at different concentrations of GAL as indicated and normalized for the change in the refractive index (ΔRI).

Adenylylation is relevant to the activity of GAL and leads to inhibition of the enzyme. For a dodecamer of GAL, an additional mass of 3.6 kDa is expected corresponding to 12-times 330 amu of adenosine-5'-monophosphate. Thus, adenylylation of GAL yields a theoretical mass of 660.7 kDa and cannot be excluded from the SEC-MALS result.

Activity of glutamate-ammonia ligase

The activity of GAL was determined by a microtiter plate assay, which was described recently by Gawronski and Benson (*158*). This assay monitors the release of inorganic phosphate that follows the hydrolysis of ATP during the forward reaction of GAL. An acidic solution of ammonium molybdate is then used to induce formation of an anionic phosphorous molybdate complex. Subsequent reduction with ascorbic acid yields molybdenum blue that correlates linearly with the amount of released phosphate. Molybdenum blue is quantified by absorption measurement at 655 nm.

Using this assay, the Michaelis-Menten parameters were determined for the conversion of L-glutamate to L-glutamine at 37 °C in a reaction mixture containing 100 mM MOPS (pH 7.5), 50 mM $MgCl_2$, 50 mM NH_4Cl and 7.5 mM ATP. First, reference kinetic parameters were obtained for commercial GAL extracted from *E. coli* (EC-GAL, Figure 5.7A). The determined K_M of 12.71 mM (Table 5.4) corresponds to a significant increase compared to typically reported K_M values for EC-GAL ranging from 2.7 to 5 mM (*159*, *160*). This hints on a reduction in the affinity for L-glutamate and may indicate inhibition of the enzyme.

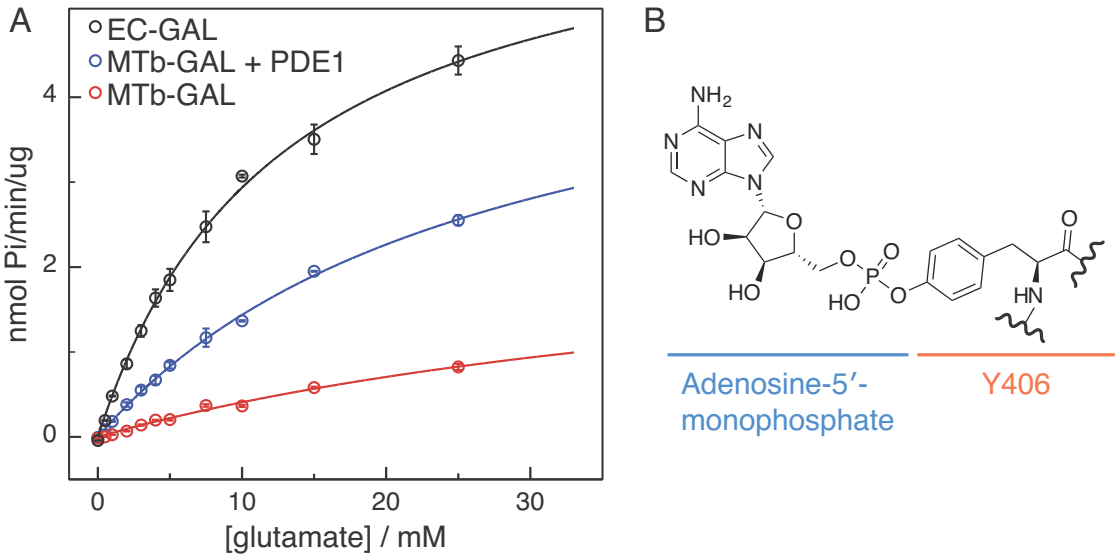


Figure 5.7 **A** Activity of glutamine-ammonia ligase (GAL). Recombinant GAL from *Mycobacterium tubercolusis* (MTb-GAL) was compared to a commercial reference obtained from *E. coli* (EC-GAL, Sigma) and MTb-GAL treated with phosphodiesterase 1 (PDE1) to remove adenylylation (MTb-GAL + PDE1). Reaction kinetics were measured in 100 mM MOPS (pH 7.5), 50 mM $MgCl_2$, 50 mM NH_4Cl and 7.5 mM ATP by varying the concentration of glutamate from 0 to 25 mM. The release of inorganic phosphate per minute normalized for the amount of GAL in the reaction was measured as described in the text. Solid lines represent individual fits according to the Michaelis-Menten kinetic equation and the corresponding Michaelis-Menten parameters are given in Table 5.4. **B** Adenylylation (AMPylation) of L-tyrosine as seen for Y406 in GAL. Adenosine-5-monophosphate forms a phosphodiester bond with the hydroxyl group of the L-tyrosine side chain.

One possible explanation for the observed decrease in the substrate affinity of EC-GAL may be the pre-incubation of the plate at 37 °C prior to the reaction. This may result in the degradation of ATP and thus the formation of ADP, which is complexed by magnesium and is reported to inhibit GAL in a competitive mode (*159*). Given the supplier's information on the activity of EC-GAL under similar conditions (pH 7.1 and 37 °C), a V_{max} of about 33 nmol Pi/min/U is expected. The obtained V_{max} of 6.67 nmol Pi/min/µg corresponds to 24 nmol Pi/min/U indicating a reduction in the rate of the reaction by about 37 %. A simultaneous decrease in the apparent V_{max} and in the apparent affinity for L-glutamate (increase in K_M) indicates mixed inhibition of EC-GAL.

The recombinantly expressed GAL from *Mycobacterium tuberculosis* (MTb-GAL, Figure 5.7A) displayed a 4.3-fold higher K_M relative to EC-GAL and in addition, a significant reduction in the V_{max} by a factor of 2.5 (Table 5.4). This apparent decrease of the substrate affinity and the maximum reaction rate again indicates mixed inhibition of MTb-GAL (*161*). Interestingly, MTb-GAL has been reported to display reduced affinity for ATP compared to native EC-GAL even in the absence of adenylylation (*157*), which possibly hints on an overall difference in the enzymatic activities.

Table 5.4 Michaelis-Menten paramaters of glutamate-ammonia ligase (GAL) for the substrate glutamate. Reaction kinetics were obtained in containing 100 mM MOPS (pH 7.5), 50 mM $MgCl_2$, 50 mM NH_4Cl and 7.5 mM ATP by varying the concentration of glutamate from 0 to 25 mM.

	V_{max} [nmol Pi/min/µg]	K_M [mM]
EC-GAL	6.67	12.71
MTb-GAL	2.65	55.05
MTb-GAL+PDE1	5.36	27.32

Since adenylylation of GAL was not ruled out by the SEC-MALS measurements, GAL was treated with phosphodiesterase 1 (PDE1), which is known to remove adenylylation (*162*). Remarkably, incubation of GAL with PDE1 (MTb-GAL + PDE1, Figure 5.7A) resulted in a two-fold increase of V_{max} from 2.65 to 5.36 nmol Pi/min/µg (Table 5.4). At the same time the apparent K_M decreased from 55.05 to 27.32 mM. Incubation with PDE1 may remove AMP from Y406 (Figure 5.7B), thereby restoring the activity of MTb-GAL. Mass spectroscopic analysis is required to verify the adenylylation state but has been impeded so far by unsuccessful recovery of the peptide comprising Y406 following tryptic digestion. However, adenylylation of MTb-GAL was previously observed for recombinant expression in *E. coli* (*157*) and may thus explain the improvement of the kinetic parameters after the treatment with PDE1.

Cost analysis of enzymatic production of L-glutamine

An estimate of the cost associated with the enzymatic synthesis of isotope-labeled L-glutamine is given in Table 5.5 based on the current market prices. Combination of the inexpensive precursors $^{15}NH_4Cl$ and 2-oxoglutaric acid with recombinant production of GAL enables the synthesis of 1 g $^{15}N_2$-L-glutamine for a total cost of 59 EUR. $^{15}N_2$-L-glutamine contributes significantly to the costs associated with the described isotope labeling media based on labeled yeast and algal extracts (Chapter 2.1 and Chapter 4). As indicated in

Table 5.3, this cost will eventually be reduced by more than ten times when $^{15}N_2$-L-glutamine is produced according to the described protocol. It can also be noted, that the additional costs related to fractional deuteration of the H^{β} positions are negligible.

Furthermore, deuteration of the H^{α} position can readily be achieved enabling the synthesis of $^{15}N_2$-L-glutamine-2,3,3'-d_3, when formic-*d* acid is used to regenerate NAD^+ in the first step of the reaction. This will increase the total costs of the procedure from 63 to 72 EUR. Although this labeling scheme is not commercially available, it is of great importance for backbone and side chain correlation experiments by NMR of larger proteins (*126*). Complete deuteration of L-glutamine requires the commercial precursor 2-oxoglutaric-3,3',4,4'-d_4 and addition of formic-*d* acid in the initial reaction step, which significantly increases the costs of the procedure. However, the enzymatic synthesis of $^{2}H_5$,$^{15}N_2$-L-glutamine is still more economical than the commercially available $^{2}H_5$,$^{15}N_2$-L-glutamine at a 2.5 times higher price (2500 EUR/g, Cambridge Isotope Laboratories).

Table 5.5 Cost estimates for enzymatic synthesis of labeled L-glutamine based on a yield of 88 %[a]

Compound (Amount)		**Labeling of L-glutamine** (1 g)		
		^{15}N	$^{15}N,^{2}H_2$[b] (H^{β})	$^{15}N,^{2}H_4$[c] (H^{β} and H^{γ})
Synthesis of L-glutamic acid				
GDH	156 U[d]	6	6	6
FDH	8 U	11	11	11
NAD^+	0.26 g	5	5	5
Na-formate	1.1 g	0.1	0.1	0.1
2-oxoglutaric acid	1.14 g	0.5	0.5	910
$^{15}NH_4Cl$	0.42 g	6	6	6
D_2O	12 ml	-	4	-
Synthesis of L-glutamine				
GAL	300 U	(382)[e]	(382)[e]	(382)[e]
$^{15}NH_4Cl$	0.36 g	5	5	5
ATP-2Na	6.6 g	25	25	25
$MgCl_2$	1.9 g	0.1	0.1	0.1
Total		59 (441)	63 (445)	968 (1350)

[a] All costs are given in Euro using conservative estimates of the market prices

[b] Synthesis of $^{15}N_2$-L-glutamine-3,3'-d_2 (H^{β}) using the homemade precursor 2-oxoglutaric-3,3'-d_2 acid

[c] Synthesis of $^{15}N_2$-L-glutamine-3,3',4,4'-d_4 (H^{β} and H^{γ}) using commercial 2-oxoglutaric-3,3',4,4'-d_4 acid

[d] Unit definition according to supplier (see Materials and Methods)

[e] Costs account for purchase of commercial GAL and may be omitted by recombinant expression (see main text)

Conclusion and perspective

A step-by-step protocol for the enzymatic synthesis of isotope-labeled L-glutamine starting with 2-oxoglutaric acid has been described. Both, high conversion yield to glutamine (>88 %) and high levels of ^{15}N and ^{2}H incorporation were obtained. In particular, ^{15}N incorporation in combination with selective deuteration of the H^{β} and H^{γ} positions was achieved. While deuteration of the H^{β} positions is performed by a base-catalyzed exchange in D_2O prior to the enzymatic reaction, additional deuteration of the H^{γ} positions is primarily established for the commercially available precursor (2-oxoglutaric-3,3',4,4'-d_4 acid). A straightforward expansion of the described protocol may provide the possibility to deuterate the H^{α} position in L-glutamine by addition of formic-*d* acid in the initial reaction step.

Furthermore, a protocol for recombinant production of glutamate-ammonia ligase (GAL) in *E. coli* was described. Additional refinement of expression conditions to increase the yield of functional GAL is still required. One possible approach could rely on minimal media that are synthetically defined and do not contain glutamine, since presence of glutamine in the medium is reported to induce adenylylation of GAL (*157*). In parallel, treatment of purified GAL with PDE1 may be routinely employed to overcome adenylylation. The ability to produce sufficient amounts of this enzyme dramatically reduces the costs of the developed method. Along these lines, recombinant expression of formate dehydrogenase and glutamate dehydrogenase in *E. coli* were recently reported (*163*, *164*). Hence, all enzymes can be obtained in their functional form by recombinant production at low overall costs. This is especially relevant for large-scale preparations of isotope-labeled glutamine. It is hoped, that this protocol including the homemade production of the required enzymes will be implemented in a standard laboratory environment to manufacture isotope-labeled glutamine in an economic manner.

Material and Methods

Repetitive batch synthesis of L-glutamic acid

Glutamate dehydrogenase from bovine liver and formate dehydrogenase from *Candida boidinii* were purchased from a commercial supplier (Roche). Detailed protocols for the synthesis and the crystallization of L-glutamic acid are provided in Table 5.1.

Polarimetric analysis of L-glutamic acid formation

Optical activity of the reaction was determined at 589 nm (Sodium D line) in a cell of 10 cm path length using a P-2000 polarimeter (Jasco Analytical Instruments). All measurements were carried out in the reaction mixture (see Table 5.1) at 25 °C.

Enzymatic synthesis of L-glutamine by glutamate-ammonia ligase

Glutamate-ammonia ligase from *E. coli* was purchased from a commercial supplier (Sigma-Aldrich). A detailed protocol of the reaction is given in Table 5.1.

Thin-layer chromatographic analysis of amino acids

Samples of the reaction mix were applied to aluminum supported silica gel TLC plates (Sigma-Aldrich). TLC plates were developed in 20:5:1 isopropanol/water/formic acid. Air-dried plates were stained with 0.1 % ninhydrin in ethanol and subsequently heated for 10 min at 100 °C.

Selective deuteration of 2-oxoglutaric acid

Preparation of 2-oxoglutaric-3,3'-d_2 acid by proton-deuterium exchange is described in Table 5.2. The deuterated precursor 2-oxoglutaric-3,3',4,4'-d_4 acid was purchased from Cambridge Isotope Laboratories.

NMR sample preparation

A protonated reference and partially deuterated samples of L-glutamic acid were prepared in 99.96 % D_2O containing 25 mM acetic-d_3 acid-*d* (99 atom % D, Sigma-Aldrich) and 100 µM 3-(trimethylsilyl)-2,2',3,3'-tetradeuteropropionic acid (TSP-d_4) at pD 5.0.

NMR spectroscopy and data processing

NMR experiments were recorded on a Bruker Avance III 600 MHz spectrometer equipped with a broadband inverse (BBI) room temperature probe. All experiments on glutamic acid were acquired at a temperature of 30 °C. Fractional deuteration of glutamic acid was analyzed by 1D ^{1}H NMR experiments using a single pulse-acquire sequence with acquisition times of 3 s. Interscan delays were optimized according to proton T_1 times obtained by standard inversion recovery experiments. NMR data were processed and analyzed with the TopSpin 3.1 software package (Bruker).

Recombinant expression and purification of glutamate-ammonia ligase in E. coli

A synthesized gene of glutamate-ammonia ligase (*glnA1)* from *Mycobacterium tuberculosis* (Uniprot entry P9WN39) was purchased from GenScript. The sequence was codon-optimized for expression in *E. coli* and an N-terminal hexahistidine-tag was added to yield the final construct MGHHHHHHGS-*glnA1*(M^1-V^{478}) (GAL). Recombinant GAL was cloned into the pET-11a vector from Millipore. Shaker flask cultures of *E. coli* BL21 cells transformed with recombinant GAL were grown at 37 °C in LB medium until OD_{600} 0.5. Protein expression was induced by addition of 1 mM isopropyl-D-thiogalactopyranoside (IPTG) and carried out for 8 h at 30 °C. Cells were harvested by centrifugation at 5000*g* for 20 min. The pellet was resuspended in ice-cold 50 mM TRIS-HCl (pH 8), 100 mM NaCl, 20 mM $MgCl_2$, EDTA-free protease inhibitor cocktail (Complete, Roche) and 20 mM imidazole (Buffer A). Cell lysates were prepared by a french pressure cell press. The supernatant was separated from the cell debris by centrifugation at 10000*g* for 20 min at 4 °C and subsequently applied to a 5 ml HisTrap FF column (GE Healthcare) equilibrated in Buffer A. Unbound sample was removed with an additional 2 column volumes (CV) of Buffer A and the resin was washed with 10 CV of 50 mM TRIS-HCl (pH 8), 100 mM NaCl, 20 mM $MgCl_2$ and 50 mM imidazole. Finally, bound GAL was eluted with 50 mM TRIS-HCl (pH 8), 100 mM NaCl, 20 mM $MgCl_2$ and 250 mM imidazole. Fractions containing GAL were pooled and concentrated in a 30 kDa MWCO centrifugal filter unit (Milipore). Concentrated GAL was applied to a HiLoad 16/600 Superdex 200 gel-filtration column (GE Healthcare) equilibrated with 20 mM TRIS-HCl (pH 7.5), 100 mM NaCl, and 50 mM $MgCl_2$.

Activity assay and de-adenylylation of glutamate-ammonia ligase

A microtiter plate setup for activity measurements of glutamate-ammonia ligase (GAL) was adapted from a previously described protocol (*158*). Briefly, substrate kinetics for glutamate were obtained in a reaction mixture containing 100 mM MOPS (pH 7.5), 50 mM $MgCl_2$, 50 mM NH_4Cl and 7.5 mM ATP by varying the concentration of glutamate from 0 to 25 mM. All measurements were carried out for 5 min at 37 °C and pH 7.5. Non-linear fitting to the Michaelis-Menten kinetic equation was determined by a routine written in Matlab (MathWorks).

Removal of adenylylation was achieved by treatment of GAL with phosphodiesterase 1 (PDE1) purified from snake venom (Sigma-Aldrich). Sufficient amounts of GAL were incubated with 5-10 U of PDE1 per mg of GAL at 37 °C for 2 h. Subsequently, PDE1 was removed by either gel-filtration chromatography on a Superdex 200 gel-filtration column (GE Healthcare) or centrifugal concentration using a 100 kDa MWCO.

6 Yeast extracts as economical source of isotope-labeled amino acids for cell-free protein synthesis

Introduction

Cell-free expression (CFE) has seen routine application for the production of isotope labeled proteins in NMR structural studies(*101*, *165-168*). Over the past two decades a variety of protocols for selective and combinatorial labeling in the CFE system have been established. Figure 6.1 depicts the principle of the continuous-exchange cell-free reaction representing the open nature of the system. CFE enables direct addition of labeled amino acids based on a wide array of isotope-labeling schemes (*169*). Furthermore, these labeling protocols greatly benefit from diminished metabolic scrambling as compared to cell-based expression (*101*, *165*). Conventional uniform labeling requires supplementation of all 20 canonical amino acids in their isotope-labeled form. Although commercial suppliers offer tailor-made amino acid mixtures for CFE systems, the respective costs of these mixtures are still high. Therefore, alternative sources of labeled amino acids are desirable. Homemade yeast extracts (YE) have been previously applied as an economical source of amino acids for isotope labeling in insect cell cultures (Chapter 2.1).

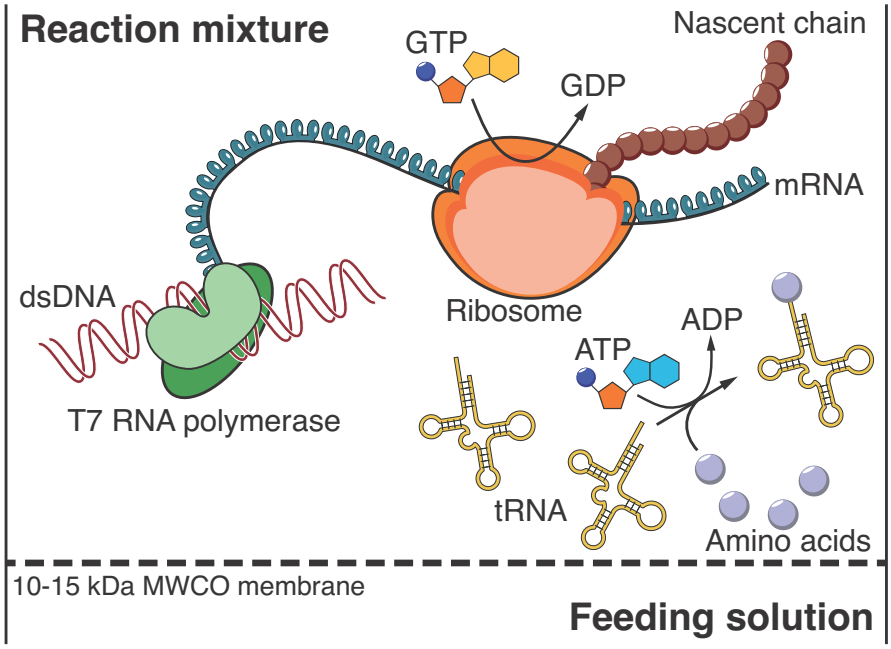


Figure 6.1 Continuous-exchange cell-free protein production. The reaction mixture containing the translation machinery is derived from an extract of cellular origin such as *E. coli*. In addition, the T7 RNA polymerase is used for transcription of the construct of interest. A dialysis membrane to enable constant exchange of the required substrates and efficient regeneration of ATP and GTP separates a feeding solution providing all reactants, which are consumed during the reaction.

This chapter describes an initial approach to employ homemade YE as the source of labeled amino acids for the *E. coli*-based CFE system before thorough analysis of the YE amino acid content had been established (Chapter 2.1). Sufficient expression levels for structural studies were obtained using YE and by additional supplementation of the amino acid cysteine. The method was then applied for uniform ^{2}H,^{15}N-labeling of the green fluorescent protein GFP providing a well-resolved ^{1}H,^{15}N-HSQC spectrum. Based on the later

obtained results of single amino acid analysis (Chapter 2), a further optimization of the supplementation scheme is suggested.

Remark on author contribution

The work of this chapter was carried out together with Dr. Sina Reckel (formerly member of the lab of Prof. Sebastian Hiller, now EPFL Lausanne), who performed all cell-free reactions and provided all additionally required reactants.

Results and Discussion

Performance of yeast extract in cell-free protein synthesis

Typically, all canonical amino acids required for protein synthesis are directly added to the cell-free reaction mixture. To test yeast extract (YE) as a source for these amino acids, an analytical scale continuous exchange cell-free expression (CECF) setup was used to express the green-fluorescent protein (GFP). GFP was chosen as a model system due to high productivity and the possibility of direct quantification by the intrinsic fluorescence of the protein.

Samples from overnight CECF runs were taken and the fraction of insoluble protein was separated by centrifugation. Afterwards, GFP concentration in the obtained supernatants was quantified by the fluorescence at 515 nm following excitation at 485 nm using a microtiter plate reader. Figure 6.2 shows the resulting protein yields for different amino acid sources. Reference expression was established for the twenty canonical amino acids at concentrations of 0.5 mM yielding 216 (± 9) µg/ml of folded GFP (Figure 6.2, 20 aa). When yeast extract was used in the CECF reaction at a concentration of 5 mg/ml an average yield of 88 (± 1) µg/ml was obtained (Figure 6.2, YE). This corresponds to a significant decrease in the expression level by about 59.3 %. Remarkably, the yield of GFP stayed constant with increasing concentrations of employed YE. This hints at a complete abscence rather than an insufficient concentration of a single compound.

In addition, algal extract (Figure 6.2, AE), which serves as the commercial source for 16 out of 20 isotope labeled amino acids, was supplemented to CECF reactions at 5, 10, and 15 mg/ml. When AE was supplemented at 5 mg/ml only 0.44 (± 0.1) µg/ml of folded GFP was obtained, while a maximum yield of 0.93 (± 0.2) µg/ml was achieved for 15 mg/ml AE. The average expression level for all AE conditions tested was equal to 0.65 µg/ml corresponding to an average reduction of about 129-fold relative to all analyzed YE conditions. Whether higher concentrations of AE would further improve the expression level was not tested at this point. However, as the overall expression was dramatically lower than the reference level, a substantial increase in the AE concentration would be required to recover the yield of GFP. Moreover, a correlation was found between the productivity and the employed concentration, which may indicate insufficient levels of several amino acids. As discussed in Chapter 4, the overall amino acid content of YE and AE was found to be similar, yet both extracts differ in their respective composition. Algal cell extracts do not contain tryptophan, cysteine, asparagine and glutamine as a result of acid hydrolysis. In contrast, all twenty amino acids can be detected in yeast extracts that are produced by autolysis of the yeast cells. Therefore, the obtained difference in GFP expression for AE and YE may simply relate to the difference in amino acid composition.

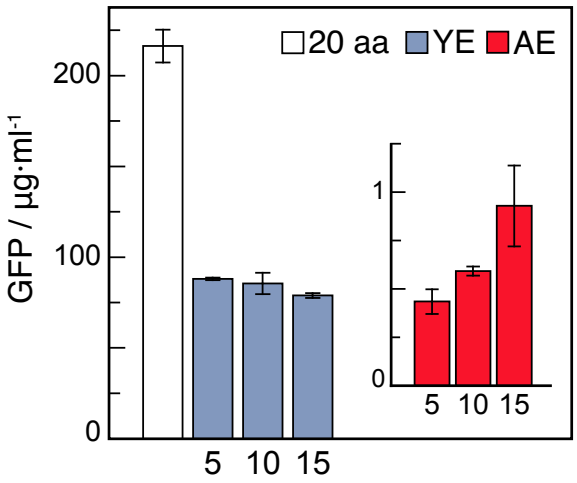


Figure 6.2 Continuous exchange cell-free production of GFP based on different amino acid sources. Expression levels were determined for a reference containing all 20 canonical amino acids (20 aa), yeast extract (YE) and algal extract (AE) shown in the inset. The concentration of GFP in the reaction mixture was determined from the fluorescence read-out of the soluble fraction, which was calibrated for GFP samples of known concentration using linear regression. Numbers below the bars indicate the concentration of the respective extract in mg/ml.

Initial combinatorial trials in insect cells showed that when YE was the only source of amino acids in the labeling media, a significant improvement of the expression yield was obtained by additional supplementation of the amino acids cysteine, aspartate, asparagine, proline and tryptophan. This was obtained by screening for changes in the expression level using SDS-PAGE analysis, since a quantitative amino acid analysis of the extracts had not been established at this point. Accordingly, it was tested whether addition of these five amino acids (CDNPW) to the YE-based CECF reaction shows a similar improvement. Figure 6.3 shows the expression level of GFP obtained when using the addition of two different batches of homemade YE (YE 1 and YE 2) at a concentration of 10 mg/ml in comparison with the addition of twenty canonical amino acids 0.5 mM each (20 aa). As before, a reduction of the expression level was observed for both YE preparations with 149 (± 17) µg/ml and 107 (± 4) µg/ml of GFP for YE 1 and YE 2, respectively. This corresponds to a relative decrease in expression yield to 55.6 % and 42.9 % when compared to the 20 aa reaction for which 255 (± 9) µg/ml of GFP were obtained. Furthermore, the data indicate a difference between the two extracts, which may relate to variations in the extract preparation. When 0.5 mM of CDNPW was supplemented to the YE-based reactions (+), recovery of GFP expression to 81.7 % and 68.8 % was observed for YE 1 and YE 2, respectively. This indicated that at least one of the added amino acids contributes significantly to the expression yield and presents a limiting factor in the prepared YE.

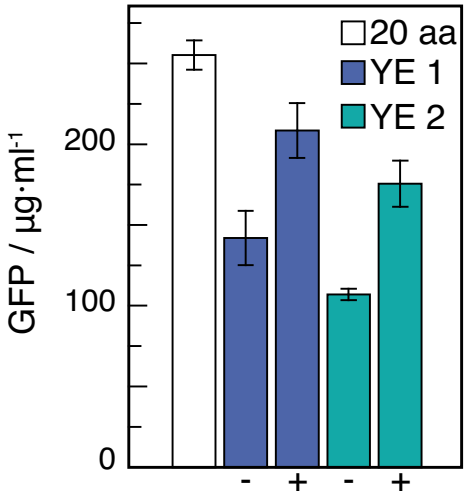


Figure 6.3 Performance of yeast extracts for cell-free production of GFP. Expression levels of GFP are shown for a reference reaction containing all 20 canonical amino acids (20 aa) and two batches of yeast extract (YE 1 and YE 2) employed at 10 mg/ml. Reactions based on YE as the sole amino acid source are denoted by the "-", while supplementation (0.5 mM each) of the amino acids C, D, N, P, and W is indicated by the "+". The concentration of GFP in the reaction mixture was quantified as described in the text.

Production of $^{2}H,^{15}N$-GFP for NMR studies

In a next step, labeling of GFP by ^{15}N and ^{2}H was investigated based on this protocol for the CECF system. To this purpose $^{2}H,^{15}N$-labeled YE was prepared as described in Chapter 2.3 and used in a CECF reaction supplemented with unlabeled CDNPW. Folded ^{15}N-GFP was obtained from the supernatant of the reaction using Ni-NTA affinity chromatography followed by a buffer exchange step for NMR spectroscopy. Figure 6.4 shows the obtained ^{1}H-^{15}N HSQC spectrum recorded for 9 h on the 60 µM sample of $^{2}H,^{15}N$-GFP. Although the NMR data are of preliminary nature, the spectrum appears well dispersed and homogenous. In total, 205 peaks were observed, which correspond to 85 % of all 241 non-proline residues. The fact that CDNPW are unlabeled reduces the number of expected residues to 207, which agrees well with the observed number of resonances. In line with this expectation, a preliminary examination indicated that the readily assignable cysteine (C48, C76) and aspartate residues (D19, D102, D197) were missing. Careful analysis of the peak intensities may still reveal differences for the CECF and the insect cell derived protein (see Chapter 2.3) regarding label dilution. A recent report indicates no detectable isotope scrambling of C, P and W (*101*). However, significant scrambling for D and N was observed. Further analysis of the isotope incorporation by mass spectroscopy (Chapter 2.2) is required to characterize isotopic scrambling as well as label dilution present in YE-based CECF reaction.

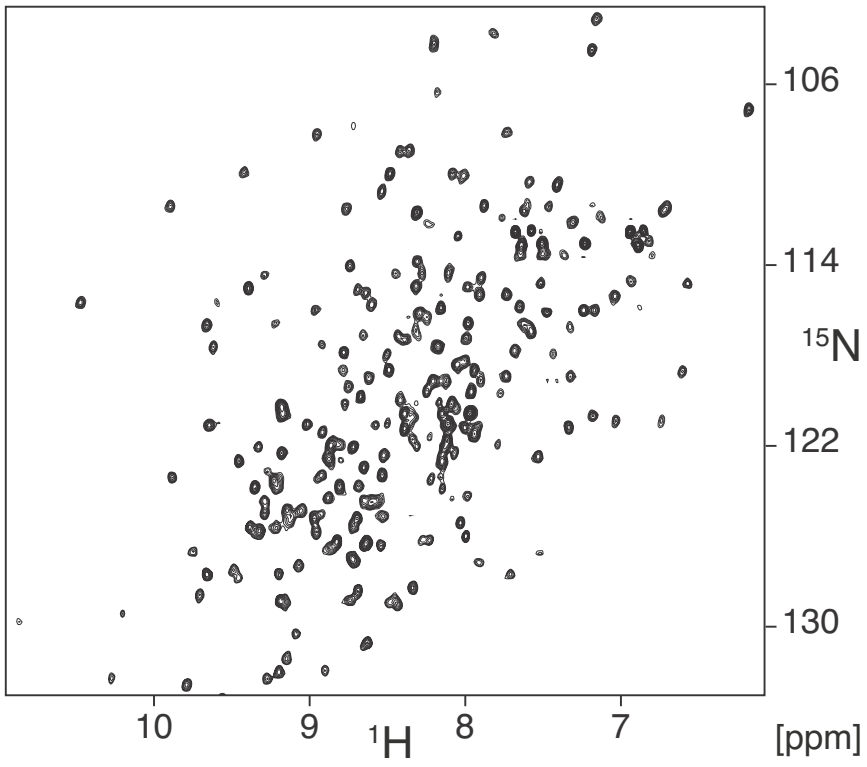


Figure 6.4 ^{1}H-^{15}N HSQC spectrum of ^{2}H,^{15}N-GFP produced in the CECF system using ^{15}N-labeled YE and unlabeled CDNPW. The spectrum was recorded in 9 h on a 60 µM sample of ^{2}H,^{15}N-GFP obtained by affinity purification as described in the text.

Refinement of the amino acid supplement

To test whether all or only a subset of the initially supplemented amino acids CDNPW is required to restore expression, each amino acid was separately removed from the supplementation scheme and the expression level of GFP was quantified. Figure 6.4 shows the resulting productivity for different CDNPW-derived combinations of amino acids in CECF reactions containing 10 mg/ml of YE. Reference expression levels were again established for supplementation of CDNPW and YE (Figure 6.5, condition 1). The overall expression yield was slightly increased compared to the previous results, where the identical batch of YE was used. Batch to batch variations in the cell-free reactions may explain this effect (S. Reckel, personal communication).

In case of YE alone (Figure 6.5, condition 8), a decrease in expression to about 48.8 % relative to the reaction supplemented with CDNPW was observed, which is in good agreement with the previous results. Remarkably, comparable expression levels were obtained when D, N, W and P where eliminated from the supplement (Figure 6.5, conditions 2-5), suggesting that none of these amino acids is essential to improve the productivity of YE. In line with this observation, supplementation of DNPW was not sufficient to restore the yield (Figure 6.5, condition 7) and displayed equally low yields of GFP as YE alone (Figure 6.5, condition 8). Therefore, supplementation was reduced to C and W recovering the yield of GFP (Figure 6.5, condition 6). Both results indicate that C is strictly required to achieve high-level expression in the YE-based CECF setup. Presumably, supplementation of W is not

required as indicated by condition 4 in Figure 6.5. It is thus expected that supplementation of C alone in addition to YE is efficient for cell-free protein synthesis.

However, supplementation of C alone was not tested in these preliminary experiments. As discussed in Chapter 2, cysteine is among the low-abundant amino acids in YE prepared by autolysis. The recovery of the expression level observed for supplementation by C-containing amino acid mixtures corroborates this finding. Although biosynthesis of cysteine is present in *E. coli* (*170*) the obtained data suggest that the respective pathway is inactive in the cell-free extract. A possible explanation for this inactivity may be given by the degradation or loss of the involved enzymes, which otherwise display significant aminotransferase and transamidation activity (*101*). Reduced metabolic scrambling of amino acids is indeed described for the cell-free system in the literature (*101*, *165*).

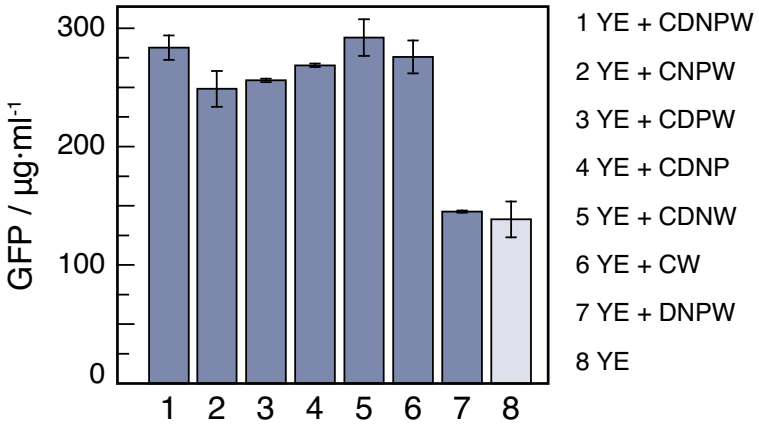


Figure 6.5 Influence of single amino acids on the productivity of yeast extract in the CECF reaction. Different combinations of the amino acids C, D, N, P, and W were added at concentrations of 0.5 mM to cell-free reactions supplemented with 10 mg/ml of YE. The expression level of GFP for individual reactions was determined as described in the text.

Rational optimization of the supplementation scheme by amino acid analysis

The experiments described above on amino acid substitution in YE-based CECF reactions were carried out before a thorough analysis of the amino acid content had been established. Using the later obtained results on the amino acid composition of YE, the limitation by individual amino acids can be quantified (Chapter 2). Figure 6.6 shows the amino acid content of CECF reactions supplemented with 10 mg/ml of YE (green bars) in comparison to concentrations of amino acid typically used for these reactions and the amino acid distribution of GFP (blue). Whereas in most cases, the concentration of amino acids exceeded those of the reference (0.5 mM, dashed line), concentrations of glutamine, histidine, methionine, and tryptophan were significantly lower in the YE-based reaction. However, supplementation of Q, H and M was not required to restore expression levels for all tested conditions (Figure 6.5). In contrast, supplementation of cysteine was strictly required but the respective concentration of C in YE has not been determined by amino acid analysis.

The limitation in glutamine is possibly compensated by activity of glutamine synthetase, which performs the conversion of glutamate to glutamine and was reported to be present in the cell-free system derived from *E. coli* (*101*) . Moreover, the concentration of histidine in the YE-based CECF reaction differed only by a factor of 1.4 and may still support high-level

expression of GFP. Despite the 2.5-fold lower concentration of methionine in YE, none of the tested supplementation schemes containing cysteine (Figure 6.5, conditions 1-6) displayed a significant decrease in overall expression compared to the reference using 20 amino acids (Figure 6.3). This indicates that in contrast to cysteine, supplementation of methionine was not essential to restore expression. S30 extracts used for CECF reactions are prepared from *E. coli* K-12 strain A19 engineered for methionine prototrophy (*171*). Therefore, cysteine may donate sulfur for biosynthesis of methionine.

Whereas the amino acid analysis suggests a possible limitation of tryptophan, no decrease in the expression level was observed when tryptophan was removed from the supplementation scheme (Figure 6.5, condition 4). Remarkably, GFP contains only a single tryptophan residue and hence, limitations of individual amino acids are expected to correlate with their abundance in the sequence of the expressed protein (Figure 6.6). Indeed, supplementation of the YE may be optimized according to the distribution of amino acids in the protein. In order to study the effect of glutamine, histidine, and methionine supplementation additional experiments are needed, but have not been carried out in this thesis due to time constraints.

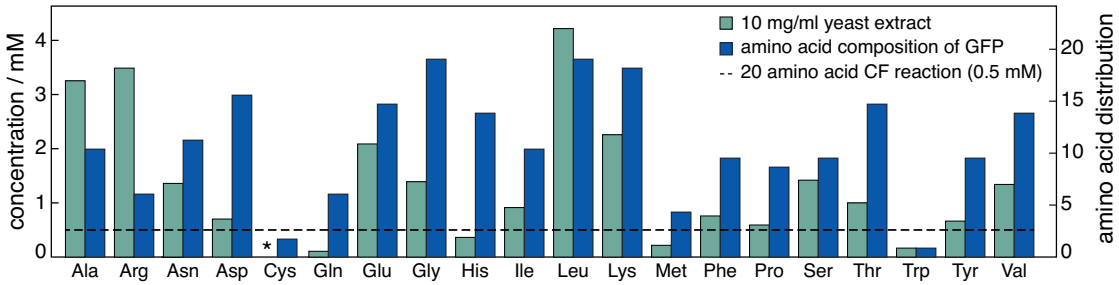


Figure 6.6 Concentrations of single amino acids in CECF reactions supplemented with 10 mg/ml of yeast extract (green) and amino acid distribution in GFP (blue). Reference concentrations of 0.5 mM used for cell-free protein synthesis based on 20 amino acid are indicated by the dashed line (*172*). The free amino acid content of the yeast extract was quantified by HPLC analysis following pre-column derivatization with phenyl isothiocyanate (Chapter 2.1). Levels of cysteine were not determined by this method (indicated by the asterisk).

Preliminary cost analysis

Given a working concentration of 10 mg/ml, 180 mg of labeled YE are required for a typical preparative scale CECF reaction volume of 18 ml (*173*). The associated costs for labeling by ^{2}H and ^{15}N are less than 10 EUR for a single reaction based on the cost analysis in Chapter 2.1 (Table 4). As shown in Figure 6.2, the concentration of YE may possibly be reduced to 5 mg/ml decreasing the reaction costs twofold. Furthermore, the obtained data supports unlabeled supplementation of cysteine for all protocols involving uniform isotope labeling and thus incurs practically no additional expense.

Commercially available cell-free amino acids mixtures contain all 20 amino acids at defined concentrations (Sigma, Cambridge Isotope Laboratories). These mixtures are derived from algal extracts and thus require addition of the four deficient amino acids C, N, Q, and W (Chapter 5) increasing the costs of the product. Based on a conservative estimate of the market prices, labeling by ^{2}H and ^{15}N using such commercial mixtures amounts to approximately 180 EUR for a typical preparative scale reaction (18 ml). Therefore, reduction

of the costs per reaction by 14-fold is possible by use of the YE-based protocol already accounting for the difference in yield.

Conclusions and perspectives

Originally, the cell-free approach was applied to screen for the required supplement of amino acids in YE-based labeling media for insect cells. Consequently, all CECF reactions were performed during an early stage in the development of the YE-based approach to isotope labeling. Eventually, the preparation of YE was optimized as described in Chapter 2.1. Therefore similar experiments employing more recent batches of YE are required. Furthermore, it is expected that production yields may only be limited by the amount of individual amino acids in the YE (Chapter 2.1). The optimal concentration of YE still remains to be determined for the modified protocol that employs supplementation of cysteine and tryptophan. Additional refinement based on the amino acid analysis of YE is required to obtain an optimal supplementation scheme for robust expression. In particular, supplementation of glutamine to the CECF reaction should be explored, which has been successfully used to obtain high protein yield in YE-based culture media for insect cells (Chapter 2.1).

An initial protocol for cell-free protein synthesis employing homemade YE as the amino acid source has been developed. The associated decrease in productivity has been overcome by addition of cysteine and tryptophan, which were required to establish high levels of expression. The combined results of different supplementation schemes (Figure 6.5) suggest that presumably tryptophan can also be omitted. This approach enables economical production of labeled proteins for NMR studies in the commonly applied CECF setup and becomes particularly advantageous for the production of proteins that require deuteration.

Uniform isotope labeling by ^{2}H and ^{15}N has been demonstrated for GFP yielding a well-resolved ^{1}H-^{15}N HSQC spectrum. The protocol can be extended to ^{13}C labeling by straightforward preparation of the respectively labeled YE. Future studies need to address isotope scrambling and label dilution, which can readily be accomplished by mass spectrometric analysis as described in Chapter 2.2.

Material and Methods

Preparation of yeast extracts

Fermentation of *Pichia pastoris* was carried out as described in Chapter 2.1. Autolysis was induced by addition of 2 % (v/v) ethanol to 30 % (w/v) yeast slurries adjusted to pH 8 and performed for five days at 50 °C. Subsequently, extracts were filtered through a 10 kDa MWCO using a stirred filtration cell (Millipore). The obtained permeate was lyophilized to yield a powder of yeast autolysate (yeastolate).

GFP construct for cell-free expression

The amino acid sequence of the GFP construct (EGFP) used to evaluate the productivity of the cell-free reaction was

```
MSKGEELFTGVVPILVELDG  DVNGHKFSVSGEGEGDATYG  KLTLKFICTTGKLPVPWPTL
VTTLTYGVQCFSRYPDHMKR  HDFFKSAMPEGYVQERTISF  KDDGNYKTRAEVKFEGDTLV
```

NRIELKGIDFKEDGNILGHK LEYNYNSHNVYITADKQKNG IKANFKIRHNIEDGSVQLAD
HYQQNTPIGDGPVLLPDNHY LSTQSALSKDPNEKRDHMVL LEFVTAAGITHGMDELYKKL
AAALEHHHHHH.

An additional C-terminal hexahistidine tag was introduced to enable affinity purification of GFP for NMR studies.

Continuous-exchange cell-free synthesis of GFP

Setup of the *E.coli*-based cell-free expression system was adapted from the recently published protocol (*173*). Analytical scale reactions were performed in homemade dialysis chambers fitted to a 24-well plate format at 30 °C for 16 h. Algal amino acid extracts (ISOGRO) were purchased from Sigma-Aldrich.

Quantification of GFP expression

Productivity of the cell-free reactions was determined by measurements of the intrinsic GFP fluorescence using a 96-well format in a microtiter plate reader (BioTek Instruments). Fluorescence of the soluble fractions was quantified at 515 nm following excitation at 485 nm. Soluble fractions containing folded GFP were obtained by centrifugation of the cell-free reactions at 15000*g* for 10 min.

Purification of $^{2}H,^{15}N$-GFP for NMR spectroscopy

^{15}N-GFP was obtained from the supernatant of the cell-free reaction by centrifugation at 15000*g* for 10 min at room temperature and subsequently applied to a 1 ml HisTrap FF column (GE Healthcare) equilibrated in 20 mM sodium phosphate (pH 7.2), 100 mM NaCl, and 20 mM imidazole. Unbound sample was removed with an additional 2 column volumes (CV) of the equilibration buffer and the resin was washed with 5 CV of 20 mM sodium phosphate (pH 7.2), 100 mM NaCl, and 50 mM imidazole. Finally, bound GFP was eluted with 20 mM sodium phosphate (pH 7.2), 100 mM NaCl and 250 mM imidazole. Fractions containing GFP were pooled and buffer-exchanged using a 10 kDa MWCO centrifugal concentration unit (Millipore). The NMR sample was prepared as a volume of 270 µl containing 60 µM of $^{2}H,^{15}N$-GFP in 20 mM BIS-TRIS (pH 7) and 50 mM NaCl.

Purification of $^{2}H,^{15}N$-GFP for NMR spectroscopy and data processing

The 2D ^{1}H-^{15}N HSQC experiment on $^{2}H,^{15}N$-GFP was recorded at 25 °C with 75 (^{15}N) x 1024 (^{1}H) complex points and acquisition times of 30 ms (^{15}N) and 40 ms (^{1}H) on a Bruker Avance III 800 MHz spectrometer equipped with a triple-resonance TCI cryoprobe. NMR data were processed with the NMRPipe software package (*128*).

7 Analysis of the proteomic response to deuteration in *E. coli*

Introduction

Deuterium labeling is of major importance for NMR studies of high molecular weight proteins (>20 kDa) and protein complexes (*2*, *149*). Efficient incorporation of deuterium is possible in the prokaryote *E. coli* and lower eukaryotic systems such as yeast by growth on deuterated media with 100 % D_2O as solvent (*2*). However, higher eukaryotic cells do not tolerate such growth conditions. Accordingly, in the last two decades deuterium labeling has been performed almost exclusively and with great success in *E. coli*. Surprisingly, adaptation of *E. coli* to growth in D_2O has not been studied systematically and thus remains poorly understood. However, the understanding of the cellular processes related to survival of *E. coli* in a highly deuterated environment may help to develop methods for deuterium labeling in higher eukaryotic cells.

Interestingly, despite successful adaptation to growth in D_2O, the biomass of *E. coli* is significantly lowered and a reduction in expression yield is generally observed (*174*, *175*). Deuterium kinetic isotope effects are known to decrease steady-state enzyme kinetics in reactions involving bond cleavage and proton transfer steps (*176*, *177*). Consequently, introduction of deuterium may significantly affect enzyme-catalyzed rates and binding equilibria. Moreover, deuteration causes changes in the thermal stability of enzymes (*178*), which may alter their lifetimes and genetic regulation of their activities. A recent metabolic study by NMR utilizing ^{13}C-labeled glucose revealed differential regulation of the tricarboxylic acid cycle (TCA) in deuterated minimal media suggesting energy starvation (*175*). However, a single key factor could not be determined to improve deuterated expression in *E. coli*.

This chapter describes a proteomic approach to study the response of *E. coli* to growth in a perdeuterated minimal medium. The analysis revealed significant changes for a total of 263 proteins upon adaptation to the deuterated environment. A functional classification of the involved genes related the observed differences to generic stress response, which is similar to temperature or osmotic stress. In agreement with previous studies, the data indicate regulatory changes in the anaplerotic reactions of the TCA (*175*) and further suggest elevated activity of the electron transport chain as well as oxidative phosphorylation to generate energy. Moreover, a substantial decrease in the expression level of 30S and 50S ribosomal proteins was observed, which is the possible cause for the decrease of protein expression in deuterated minimal media.

Results and Discussion

Adaptation of E. coli to growth in a perdeuterated environment

Robust expression of uniformly deuterated proteins in *E. coli* requires one or more intermediate adaptation steps in minimal media typically comprising fractions of 30, 50 or 70 % D_2O. Growth in these adaptation cultures is limited to the end of the exponential growth phase, effectively restraining the number of cell divisions and thus the likelihood of mutations (*179*). Therefore, short-term adaptation of *E. coli* to growth in a perdeuterated minimal medium was studied on the proteome but not on the genome level.

First, cultures of the *E. coli* expression strain BL21(D3) were grown in protonated M9 minimal medium (light cultures) to the end of the exponential growth phase. These cultures were used to obtain the reference expression level in a protonated environment. In the next step, cells were adapted to growth in D_2O using M9 medium containing 70 % D_2O and finally grown in perdeuterated M9 medium containing 99.8 % D_2O and $^2H_7$-glucose (heavy cultures) to emulate the protocol used for uniform deuteration.

Growth of *E. coli* on the nutritionally rich lysogeny broth (LB) medium typically results in mean generation times of ~0.5 h. When grown on minimal media, the mean generation time increased to ~1 h in protonated M9, ~1.5 h in M9 containing 70 % D_2O, and ~2.5 h in perdeuterated M9, respectively. To obtain uniform adaptation of *E. coli* to growth in these minimal media, cultures were grown for five generations corresponding to the end of the exponential growth phase.

Light and heavy cultures were normalized for their total protein content and subject to LC/MS analysis following tryptic digestion to identify changes in the peptide level between the two cultures.

Analysis of the proteomics data

Initially, proteomics data were analyzed using the commercial software package *Progenesis IQ* (Nonlinear Dynamics, version 2.0). This approach yielded robust peak detection results for a total of 10748 light peptides, which were subsequently identified using *Mascot* (Matrix Science). However, the corresponding heavy peptides were not correctly detected and thus, quantification of these peptides was not possible. *Progenesis* employs a peak-modeling algorithm based on continuous wavelet transformation in combination with isotope pattern matching (*180*), which enables robust peak detection and handling of the noisy MS data (*181*). Whereas highly accurate peak identification was obtained for the natural abundance isotopologue pattern of the light peptide, pattern matching is impeded for the corresponding heavy peptide by complex convolution patterns of deuterium due to incomplete deuteration. More specifically, residual protonation of the solvent (0.2 %) and $^2H_7$-glucose (<2 %) results in isotopologue distributions comprising protonated isotopomers. The abundance of these isotopomers increases with the number of aliphatic proton positions for a given peptide and hence is significant for the quantification.

Quantification of the protein level in deuterated E. coli cells

To overcome the above described quantification problem, an in-house algorithm for exact detection of deuterated isotopic envelopes (EDDIE) was used to allow quantification of the heavy peptides. EDDIE relies on the prediction of the isotopic pattern based on the

incorporation level of deuterium and accounts for a shift in retention time on the reverse phase column caused by deuteration of aliphatic hydrogen positions of the amino acids.

In a first step, peak detection was verified for a total of 10748 light peptides, which were combined per protein to obtain protein level intensities (Figure 7.1A). The intensities obtained by EDDIE correlated strongly ($r^2 = 0.99$) with the results determined by *Progenesis*. This method was then used for quantification of the related heavy peptides in the perdeuterated cultures. To account for the occurrence of protonated isotopologues, quantification was performed by summation of the intensities for the five most abundant isotopologue peaks of the peptide matching the predicted m/z values. These five peaks typically present ~99.5% of the theoretical isotopologue pattern, which was calculated using a binomial distribution and a conservative estimate of the residual protonation level of 1.5 % based on deuterium labeling of individual medium components (see Materials and Methods). Unique peptide intensities were combined for individual proteins to obtain the protein level intensities. The correlation between light and heavy protein intensities ($r^2 = 0.84$) determined by EDDIE indicated that the majority of the protein expression levels were not altered as a response to growth in a deuterated environment (Figure 7.1B). This hints at a specific change in the proteome of D_2O-adapted cultures.

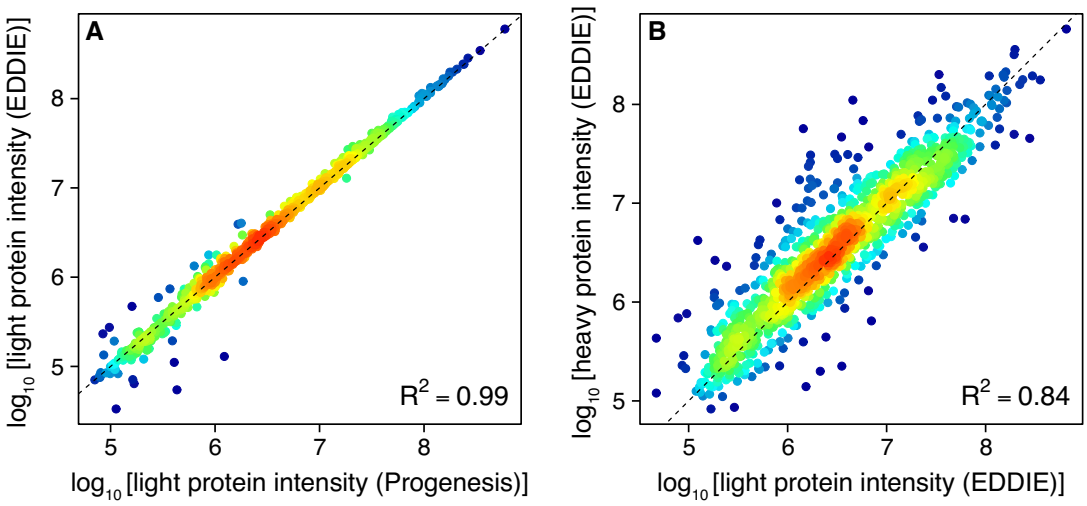


Figure 7.1 A Correlation of logarithmic protein intensities determined by *Progenesis IQ* and the in-house algorithm EDDIE. Light peptides were identified in whole cell lysates of *E. coli* cultures grown in protonated minimal medium (light cultures) following tryptic digestion. Protein level intensities were obtained by combining the annotated light peptide intensities as described in Materials and Methods. **B** Quantification of the protein intensities in perdeuterated relative to protonated *E. coli* cells. The heavy peptide intensities were quantified in whole cell lysates of *E. coli* cultures grown in perdeuterated minimal medium (heavy cultures) by EDDIE using the sum of the five most abundant deuterated peaks for a single heavy peptide (see main text) and converted to the related heavy protein intensities.

Changes in the protein expression upon growth in D_2O

Protein expression levels were analyzed for three biological replicates of *E. coli* BL21(D3) cells cultured in protonated M9 minimal medium (light cultures) and perdeuterated M9 medium containing 99.8 % D_2O and $^2H_7$-glucose as the sole carbon source (heavy cultures). Heavy/light protein ratios were calculated per protein based on quantification of the peptide

level by EDDIE to identify the change in the proteome. The obtained ratios were then tested for differential expression using moderated t-statistics based on empirical Bayes estimators (*182*) and corrected for multiple comparison using the Benjamini-Hochberg procedure (*183*). This approach estimates the false discovery rate for a change in the abundance of a single protein expressed as the q-value (see Materials and Methods).

Figure 7.2 shows a double logarithmic plot for the calculated heavy/light protein abundance ratios and the corresponding q-values. From a total of 1082 detected proteins, 420 were recognized as confident hits using a q-value cut-off of 0.01. These confident hits were further constrained using a 2-fold abundance ratio cut-off identifying 239 proteins (Figure 7.2, colored circles), for which a significant difference in the expression level was observed. Compared to *E. coli* cultures grown in protonated minimal medium, in D_2O-adapted cells the expression was lowered for 76 proteins (Figure 7.2, brown circles) and increased for 163 proteins (Figure 7.2, blue circles). This indicates that the proteomic response to high levels of deuterium is very likely not limited to only a single metabolic or regulatory pathway.

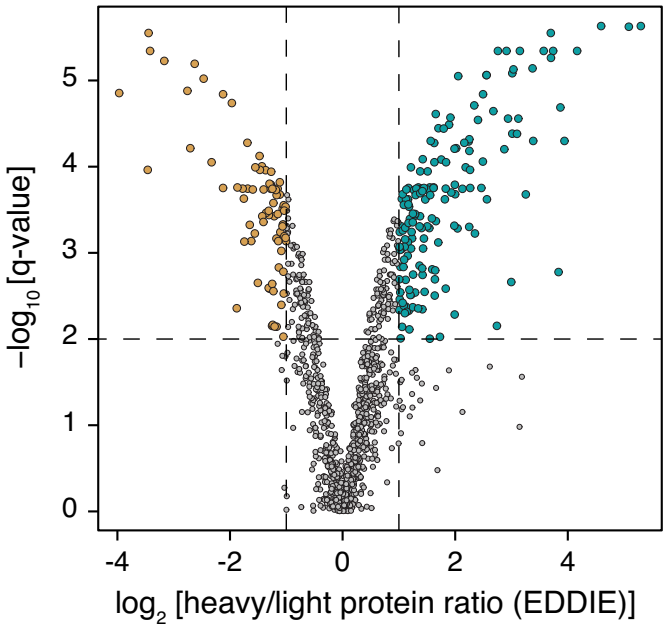


Figure 7.2 Proteome-wide response of *E. coli* to growth in a perdeuterated environment. Significant changes in the protein abundance (q-value ≤0.01) corresponding to a 2-fold increase or decrease in the expression level are shown in blue and brown, respectively. Heavy/light ratios were calculated from protein level intensities determined in whole cell lysates of *E. coli* cultures grown in protonated (light) and perdeuterated (heavy) M9 minimal medium. Protein level intensities were determined by quantification of the annotated light and corresponding heavy peptide intensities using EDDIE as described in the main text. The q-values reflect a significant change of the expression level controlled by the false discovery rate (see Materials and Methods).

It should be noted that the difference in protein abundance level could also in part be due to the change from a nutritionally rich LB to the M9 minimal medium. More specifically, the overall number of generations grown in minimal medium differs for the protonated (five generations) and perdeuterated culture (fifteen generations). To identify possible changes in the protein abundance level related to adaptation of *E. coli* to growth on protonated M9

medium alone, extended subculturing in this medium is required to obtain the same overall number of generations with respect to the perdeuterated culture. This will be verified in additional experiments. However, for the present analysis is was assumed that such an effect is insignificant as metabolic adaptation of *E. coli* predominantly occurs within the initial lag phase of the culture (*184*).

Differential expression data were grouped by their biological function to identify the response to D_2O adaptation using the functional annotation tool of the *DAVID* bioinformatics resources (http://david.ncifcrf.gov, v6.7) (*185*). Out of 236 identified, 25 genes were uncharacterized according to the functional cluster annotation table of *DAVID*. This refers to proteins, for which comprehensive information on the structure and function is limited and for some cases has been predicted by *in silico* methods alone. Consequently, careful analysis of these proteins is required to identify their biological context and their relevance for adaptation.

A comprehensive discussion of all identified proteins is beyond the scope of this thesis and thus a limited number of proteins will now be discussed in detail, which may contribute to an understanding of the adaptation to growth in high levels of D_2O.

Response to abiotic stress

Among all identified groups, the most notable change in the protein expression in the perdeuterated minimal medium was related to a response to abiotic stress (Table 7.1, abiotic stress response). Such a response is typically induced by limitation of a substrate, temperature or osmotic stress. The expression levels in this functional group were significantly increased for all genes but the cold shock protein (*cspA*).

Entericidin B (*encB*) and the osmotically-inducible lipoprotein E (*osmE*) are directly linked to osmotic stress. At high osmolarity conditions, *encB* was found to induce cell death acting as a bacteriolytic lipoprotein (*186*). Similarly, increased levels of *osmE* were found in cultures grown in a hyperosmotic medium and associated with a modification of the supercoiled DNA structure (*187*). The hyperosmotic behavior of deuterium oxide is well-known to cause shrinking of living cells (*188*). A possible explanation is the decrease in the zero-point energy of deuterium compared to hydrogen (*189*), which leads to a slightly stronger bond between the deuteron and oxygen in D_2O, thereby lowering the self-ionization tendency of the molecule. Consequently, the solubility of ions is lowered in D_2O (*190*). This may influence the activity of ion channels, which are essential to regulate the cell volume (*191*). Remarkably, the expression level of the morphogene *bolA* was increased 13-fold, which has been described to cause a round morphology (*192*). A sphere-like cell shape effectively offers to smallest surface-area-to-volume-ratio and may function to minimize the contact area with the hyperosmotic solvent. Furthermore, overexpression of three lipoproteins (*ybjP*, *yceB*, and *ydcL*) was observed, which have not been functionally characterized. As these proteins are localized in the outer membrane of *E. coli*, they may play a role in adaptation to osmotic stress.

Remarkably, the proteomic response to growth in a perdeuterated environment in part resembles a protective reaction to temperature stress in *E. coli*. Incorporation of deuterium directly affects the stability of cellular macromolecules and may cause similar structural changes as imposed by a high temperature shock. However, genes such as *endo IV* are further indicative for a physiological response to oxidative stress induced by a possible accumulation of pro-oxidants (*193*). Furthermore, increased levels of glutaredoxin-2, thiol peroxidase (*tpx*),

and the manganese-dependent superoxide dismutase (*sodA*) were observed (Table 7.1, electron transport reactions), which function to prevent cell damage by quenching of active oxygen species. Pro-oxidants are naturally occurring during aerobic energy generation (*193*) but may accumulate as a consequence of the significant changes in oxidative phosphorylation and the electron transport chain. A response to oxidative stress was further supported by upregulation of the multidrug resistance protein (*mdtE*), for which activity was shown to increase upon oxygen deprivation (*194*). The requirement for rigorous aeration to enable high-level expression in highly deuterated cultures of *E. coli* (*175*) may be rationalized by the observed increase of *mdtE*.

Interestingly, a substantial 7.7-fold increase in the protein abundance was found for the universal stress protein G and was even more pronounced for the related universal stress protein F (14.3-fold). These two genes have been reported to show an additive effect in response to oxidative stress and promote cellular adhesion (*195*).

Changes in the stability of cellular macromolecules

It is expected that incorporation of deuterium and the solvent effect of D_2O may alter the strength of non-covalent interactions, which are the key-stabilizing elements in secondary structures of proteins and nucleic acids (*178*, *189*, *196*). Therefore, genes associated with a response to abiotic stress were classified for their function in maintaining the structural integrity of the cell as well as the assembly and disassembly of cellular macromolecules. The resulting list of functional groups contained several chaperones, proteases, DNA repair enzymes, and regulatory proteins involved in stress response pathways as well as ribosomal proteins (Table 7.1), which necessarily intersect with the previously described functional groups.

Increased levels of the heat shock protein *hslJ* and the small heat shock protein *ibpA* were observed (*197*, *198*). Heat shock proteins resemble a family of ubiquitous chaperons essential to the protein folding machinery and temperature-induced stress response in many cells (*199*). Recent studies in yeast grown in D_2O and deuterium-resistant *Chlorella* mutants linked adaptation to growth in a highly deuterated environment to increased levels of the heat shock proteins 70 and 60 (*200*, *201*). However, the protein abundance level of *dnaK*, the *E. coli* homolog of the heat shock protein 70, was only changed by about 1.8-fold, albeit satisfying the false discovery rate (q-value ≤ 0.01). Therefore, *dnaK* was not considered for the present evaluation but is possibly involved in the response to deuteration. Accordingly, an 8-fold increase in the protein abundance level was observed for the thiol:disulfide interchange protein *dsbC*. Previously it was shown that *dsbC* functions as an isomerase to resolve erroneously formed disulfides in proteins (*202*). Overexpression of heat shock chaperones suggests that protein misfolding occurred more frequently and hence may require reshuffling of non-native disulfide bonds in folding intermediates of cysteine-containing proteins.

Similarly, genes specifically involved in DNA repair were significantly altered upon deuteration. Most notably, upregulation was observed for the endonuclease 4 (*endo IV*) and the *uvrABC* system protein B (*uvrB*). Both proteins are involved in the excision repair mechanism of damaged double-stranded DNA such as base loss, deamination or oxidation (*203*, *204*). This process is known as transcription-coupled DNA-repair and resembles a rate-limiting step in transcription (*205*).

Regulation of transcription

Gene functional analysis further indicated differential expression of proteins linked to regulation of transcription (Table 7.1, regulation of transcription), which were mainly upregulated. The integration host factor subunit beta (*ihfB*) is known to be involved in the maintenance of the DNA architecture, acting globally to ensure correct DNA replication and transcription (*206*). Interestingly, *ihfB* was reported to promote bending and supercoiling of DNA (*207*, *208*) and is possibly linked to increased levels of *osmE*. Interestingly, osmotic stress has been shown to increase the level of DNA supercoiling in *E. coli* (*209*), which is in line with the above described hyperosmotic behavior of D_2O and the observed response to abiotic stress. Moreover, two mRNA chaperones, the DNA-binding protein *stpA* (*210*) and the cold-shock protein *cspA* were identified. Interaction of the latter was reported to prevent formation of non-native secondary structure (*211*). Incorporation of deuterium possibly alters stability and folding of mRNA caused by a difference in the relative stabilities of hydrogen bonds in the molecule (*189*). Therefore, a change in the expression levels of *cspA* and *stpA* possibly indicates misfolding of mRNA, which is likely to interfere with transcription as well as translation.

Changes in the ribosome biogenesis and translation

A possible alteration of the translation in *E. coli* adapted to growth in a perdeuterated environment was indicated by a decrease in the abundance of several 30S and 50S ribosomal proteins (Table 7.1, translation). This hints at an overall decrease in translation in these cultures. However, this decrease may as well be interpreted as a modulation of the transcriptional activity to facilitate survival of the D_2O-induced stress, which is possibly favored over proliferation and growth at these conditions. Moreover, the 30S ribosomal protein S4 (*rpsD*) is not only a structural constituent of the ribosome, it also functions as a negative regulator of translational initiation and displays translational repressor activity (*212*, *213*). The observed decrease in *rpsD* possibly indicates a difference in the regulation of translation upon deuteration. Again, this response may be explained by changes in the stability of the different RNA species (mRNA, rRNA, and tRNA) or the ribosome itself caused by deuterium incorporation.

Furthermore, the level of the ribosome-associated inhibitor A (*raiA*), known to regulate transcription, was found to be increased 3-fold. The activity of *raiA* was described to compensate for errors during translation by counteracting miscoding (*214*), which possibly links overexpression of *raiA* to a change in the pairing of amino acids with cognate tRNA typically observed for temperature-induced stress.

In contrast to the overall reduction in the protein abundance level of ribosomal proteins, a concurrent upregulation of the expression was observed for the arginyl-tRNA synthetase (*argS*). Oxidative stress was reported to significantly reduce the translational fidelity by impairment of the aminoacyl-tRNA synthetase editing site, eventually inhibiting growth of *E. coli* in the presence of reactive oxygen species (*215*). Accordingly, high levels of D_2O are likely to cause mistranslation. This may be explained by (i) impairment of upregulated *argS* or (ii) altered stability of the corresponding aminoacyl-tRNA, and is possibly counteracted by increased expression of the related gene.

Modification of the energy metabolism

Adaptation to growth in minimal media containing high fractions of D_2O was shown to cause distinct changes in the metabolic flux of the TCA (*175*). To identify such a change on the proteome level, genes related to the cellular energy generation were grouped using the functional annotation tool of *DAVID* (Table 7.1, TCA). In general, differential expression was observed for several enzymes involved in the aerobic energy metabolism of *E. coli*. Consequently, electron transport and oxidative phosphorylation reactions may vary significantly in the perdeuterated environment, which suggests a change in the cellular energy demand (Table 7.1, electron transport reactions and oxidative phosphorylation). Most notably, an 8-fold increase in the protein level abundance was found for the hydrogenase-1 large chain. The membrane-bound hydrogenase-1 is involved in the generation of the proton motive force in the presence of oxygen and has been reported to appear sensitive to osmotic stress (*216-218*). Hence, upregulation of hydrogenase-1 is in line with the observed response to osmotic stress in D_2O-adapated cultures.

Moreover, rearrangement of the TCA was identified in accordance with the changes in the respiratory energy generation. In particular, a significant increase in the expression level of phosphoenolpyruvate carboxykinase (*pckA*) and downregulation of pyruvate dehydrogenase (*aceE*) was observed. Whereas *pckA* catalyzes the synthesis of phosphoenolpyruvate from the TCA-cycle intermediate oxaloacetate, the activity of *aceE* involves decarboxylation of pyruvate to yield acetyl-coA. Regulation of these enzymes is required for metabolic flux control of the carbon metabolism (*219*, *220*). Together, the identified change in their protein abundance levels suggests a decrease in glycolysis and a concurrent activation of gluconeogenesis (*219*).

Remarkably, expression of isocitrate lyase was upregulated 18-fold, which strongly suggests increased activity of the glyoxylate bypass (*221*). This was further supported by significant overexpression of the transcriptional regulator *kdgR* responsible for regulation of the glyoxylate shunt (*222*). The glyoxylate shunt is an alteration of the TCA used for anaplerotic replenishment of oxaloacetate and characteristic for utilization of carbon sources other than glucose (*221*). This metabolic bypass reaction resembles an important step in the gluconeogenesis pathway and avoids carbon loss due to two consecutive decarboxylation steps in the regular TCA. Rearrangement of the metabolism may indicate a severe decrease in the rate of glycolysis, which may trigger alternative pathways such as the glyoxylate shunt to fulfill the energy requirement of the cell in a highly deuterated environment. The second enzyme essential to this pathway, which resembles an anaplerotic alteration of the TCA, is malate synthase (*glcB*). Indeed, a significant increase was observed in the protein abundance level of *glcB*. However, the corresponding q-value (0.09) was larger than the cut-off ($\leq$0.01) applied in the present analysis.

Biosynthesis of tryptophan and differential expression of ABC transporters

Moreover, a functional group was identified comprising proteins involved in the biosynthesis of tryptophan (Table 7.1, tryptophan biosynthesis). The role of tryptophan for adaptation to growth in a perdeuterated environment is unclear. Conversely, biosynthesis of tryptophan is regulated by nutritional stress and activity of this pathway has been reported to increase in minimal media deficient in tryptophan (*223*). Whereas *E. coli* cultures grown in protonated minimal medium were directly transferred from LB medium, fully D_2O-adapated

cells were grown on the basis of this minimal medium for two more cultures (70 % and 99.8 % D_2O). Consequently, the protocol used for adaptation of *E. coli* may have caused the observed change in the expression level.

Differential expression was also observed for proteins belonging to the transport system superfamily of ATP-binding cassette transporters (Table 7.1, ABC transporters). However, this group has not been considered at this point, as a comprehensive analysis requires further information on the role of these transporters for adaptation to growth on minimal medium. Again, the length of the overall adaptation time in minimal medium may have affected the expression of the related transporters and thus the transport of several nutrients irrespective of the perdeuterated environment. This possibility will be investigated in further control experiments.

Table 7.1 Selected changes in the protein expression level of *E. coli* cultures adapted to growth in perdeuterated minimal media with q-values ≤0.01. Functional clusters and gene annotations were obtained using the *DAVID* bioinformatics resource.

UniProt accession	**Gene name**	**Heavy/light ratio** (log_2)	**q-value** ($-log_{10}$)
ABC transporters			
P0AG82	Phosphate-binding protein pstS	1.38	3.95
P30859	Arginine-binding periplasmic protein 1	1.01	3.05
P0AFK9	Spermidine/putrescine-binding periplasmic protein	1.11	2.3
P09551	Lysine-arginine-ornithine-binding periplasmic protein	3.87	4.69
P0AD96	Leu/Ile/Val-binding protein	5.30	5.63
P30860	Arginine-binding periplasmic protein 2	-1.06	3.33
P23843	Periplasmic oligopeptide-binding protein	-1.01	3.17
P0AEX9	Maltose-binding periplasmic protein	-1.31	3.49
P02925	D-ribose-binding periplasmic protein	1.22	2.52
P37902	Glutamate/aspartate periplasmic-binding protein	1.84	4.05
Abiotic stress			
P0A9X9	Cold shock protein cspA	-2.33	4.05
P0A6C1	Endonuclease 4	3	2.66
P0ADB7	Entericidin B	4.59	5.63
P52644	Heat shock protein hslJ	1.25	2.64
P00946	Mannose-6-phosphate isomerase	2.74	2.15
P37636	Multidrug resistance protein mdtE	1.27	3.94
P0ADB1	Osmotically-inducible lipoprotein E	2.25	4.32
P0AFH8	Osmotically-inducible protein Y	3.37	5.14
P0C0L2	Peroxiredoxin osmC	3.38	4.3
P23894	Probable protease htpX	1.56	3.67
P0ABE2	Protein bolA	3.70	5.55
P0C054	Small heat shock protein ibpA	1.37	2.85
P75818	Uncharacterized lipoprotein ybjP	1.25	4.54
P0AB26	Uncharacterized lipoprotein yceB	2.91	5.34
P64451	Uncharacterized lipoprotein ydcL	2.13	3.74
P0A8F8	UvrABC system protein B	2	3.68
Aerobic energy metabolism			
P0AFG3	2-oxoglutarate dehydrogenase E1 component	1.62	3.76

UniProt accession	Gene name	Heavy/light ratio (log_2)	q-value ($-log_{10}$)
P25516	Aconitate hydratase 1	2.76	5.34
P0A9P0	Dihydrolipoyl dehydrogenase	-1.23	3.58
P0AFG6	Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex	1.13	3.62
P0ACD8	Hydrogenase-1 large chain	3.03	5.13
P08200	Isocitrate dehydrogenase [NADP]	1.16	3.12
P0A9G6	Isocitrate lyase	4.17	5.34
P61889	Malate dehydrogenase	1.61	3.75
P00393	NADH dehydrogenase	-2.12	4.84
P0AFD1	NADH-quinone oxidoreductase subunit E	1.25	2.37
P31979	NADH-quinone oxidoreductase subunit F	1.07	3.68
P0AC41	Succinate dehydrogenase flavoprotein subunit	1.40	3.74
P0AC44	Succinate dehydrogenase hydrophobic membrane anchor subunit	2.26	3.96
P07014	Succinate dehydrogenase iron-sulfur subunit	2.24	4.29
P0AGE9	Succinyl-CoA ligase [ADP-forming] subunit alpha	1.66	4.61
DNA repair			
P0A6C1	Endonuclease 4	3	2.66
P15032	Exodeoxyribonuclease 8	-3.97	4.85
P0A7G6	Protein recA	-1.21	3.42
P30958	Transcription-repair-coupling factor	-1.17	2.14
P0A8F8	UvrABC system protein B	2	3.68
Electron transport reactions			
P0ABJ9	Cytochrome d ubiquinol oxidase subunit 1	-2.7	4.88
P0A9P0	Dihydrolipoyl dehydrogenase	-1.23	3.58
P0AC59	Glutaredoxin-2	1.20	3.39
P00393	NADH dehydrogenase	-2.12	4.84
P0AFD1	NADH-quinone oxidoreductase subunit E	1.25	2.37
P28304	Quinone oxidoreductase	1.25	3.45
P27306	Soluble pyridine nucleotide transhydrogenase	1.55	2
P0AC41	Succinate dehydrogenase flavoprotein subunit	1.4	3.74
P07014	Succinate dehydrogenase iron-sulfur subunit	2.24	4.29
P46853	Uncharacterized oxidoreductase yhhX	1.14	2.32
Osmotic stress			
P0ADB1	Osmotically-inducible lipoprotein E	2.25	4.32
P0AFH8	Osmotically-inducible protein Y	3.37	5.14
P0C0L2	Peroxiredoxin osmC	2	4.3
P0ABE2	Protein bolA	3.7	5.55
Protein folding			
P29131	Cell division protein ftsN	1.06	2.14
P0AEU7	Chaperone protein skp	-1.27	3.74
P0ABZ6	Chaperone surA	1.29	3.45
P0ACG1	DNA-binding protein stpA	1.47	3.62
P0AEX9	Maltose-binding periplasmic protein	-1.31	3.49
P23843	Periplasmic oligopeptide-binding protein	-1.01	3.17
P0AFK0	Protein pmbA	1.5	3.74
P0AEG6	Thiol:disulfide interchange protein dsbC	3.01	4.38

UniProt accession	Gene name	Heavy/light ratio (log_2)	q-value ($-log_{10}$)
Oxidative phosphorylation			
P0ABA6	ATP synthase gamma chain	1.22	2.53
P68699	ATP synthase subunit c	-3.45	5.55
P0ABJ9	Cytochrome d ubiquinol oxidase subunit 1	-2.76	4.88
P00393	NADH dehydrogenase	-2.12	4.84
P0AFD1	NADH-quinone oxidoreductase subunit E	1.25	2.37
P31979	NADH-quinone oxidoreductase subunit F	1.07	3.68
P0AC41	Succinate dehydrogenase flavoprotein subunit	1.4	3.74
P0AC44	Succinate dehydrogenase hydrophobic membrane anchor subunit	2.26	3.96
P07014	Succinate dehydrogenase iron-sulfur subunit	2.24	4.29
Regulation of transcription			
P0A7V8	30S ribosomal protein S4	-1.4	3.74
P69913	Carbon storage regulator	1.11	2.3
P0A9X9	Cold shock protein cspA	-2.33	4.05
P0ACG1	DNA-binding protein stpA	1.47	3.62
P0A6Y1	Integration host factor subunit beta	1.05	3.63
P0AD49	Ribosome-associated inhibitor A	1.55	3.76
P0C054	Small heat shock protein ibpA	1.37	2.85
P76268	Transcriptional regulator kdgR	1.25	3.35
TCA			
P0AFG3	2-oxoglutarate dehydrogenase E1 component	1.62	3.76
P25516	Aconitate hydratase 1	2.76	5.34
P0A9P0	Dihydrolipoyl dehydrogenase	-1.23	3.58
P06959	Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex	-1.69	4.28
P0AFG6	Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex	1.13	3.62
P08200	Isocitrate dehydrogenase [NADP]	1.16	3.12
P61889	Malate dehydrogenase	1.61	3.75
P22259	Phosphoenolpyruvate carboxykinase [ATP]	1.21	3.99
P0AFG8	Pyruvate dehydrogenase E1 component	-1.97	4.74
P0AC41	Succinate dehydrogenase flavoprotein subunit	1.4	3.74
P0AC44	Succinate dehydrogenase hydrophobic membrane anchor subunit	2.26	3.96
P07014	Succinate dehydrogenase iron-sulfur subunit	2.24	4.29
P0AGE9	Succinyl-CoA ligase [ADP-forming] subunit alpha	1.66	4.61
P30178	Uncharacterized oxidoreductase ybiC	1.42	4.09
Temperature shock			
P0A9X9	Cold shock protein cspA	-2.33	4.05
P52643	D-lactate dehydrogenase	-1.78	3.75
P52644	Heat shock protein hslJ	1.25	2.64
P0AG82	Phosphate-binding protein pstS	1.38	3.95
P23894	Probable protease htpX	1.56	3.67
P0A7G6	Protein recA	-1.21	3.42
P0AD49	Ribosome-associated inhibitor A	1.55	3.76
P0C054	Small heat shock protein ibpA	1.37	2.85
P39177	Universal stress protein G	2.94	4.56

UniProt accession	Gene name	Heavy/light ratio (log_2)	q-value ($-log_{10}$)
Translation			
P0A7S3	30S ribosomal protein S12	-1.51	2.65
P0A7V3	30S ribosomal protein S3	-1.12	3.68
P0A7V8	30S ribosomal protein S4	-1.4	3.74
P0ADY3	50S ribosomal protein L14	-1.11	3.24
P0A7K6	50S ribosomal protein L19	-1.15	3.45
P60422	50S ribosomal protein L2	-1.07	3.31
P0A7L3	50S ribosomal protein L20	-1.62	3.14
P0A7Q1	50S ribosomal protein L35	-1.31	3.44
P11875	Arginyl-tRNA synthetase	1.3	3.67
P0A8N5	Lysyl-tRNA synthetase, heat inducible	-1.18	3.17
P0AD49	Ribosome-associated inhibitor A	1.55	3.76
P16456	Selenide, water dikinase	1.59	2.81
P0A8L1	Seryl-tRNA synthetase	-1.17	3.67
Tryptophan biosynthesis			
P00895	Anthranilate synthase component 1	-1.13	2.83
P07639	3-dehydroquinate synthase	3.74	5.34
P00888	Phospho-2-dehydro-3-deoxyheptonate aldolase	-1.05	2.53
P00909	Tryptophan biosynthesis protein trpCF	1.04	3.31
P00904	Anthranilate synthase component II	1.35	3.76

Conclusions and perspectives

Proteomic analysis of *E. coli* adapted to growth in perdeuterated minimal medium was established to identify changes in the protein expression level. A key challenge for this approach is the adaptation procedure, which requires serial passage of *E. coli* in minimal media containing increasing levels of D_2O. In the present study, adaptation was achieved by growth of *E. coli* for five generations in the respective minimal medium to obtain uniformly adapted cultures. As cells were first grown on nutritionally rich lysogeny broth (LB) medium, a possible response may occur to growth on minimal medium and may still be apparent in the proteome of perdeuterated cultures. To identify such a response based on the described approach, changes in the protein expression level may be analyzed in cultures continuously grown on protonated minimal medium alone or using cultures grown on LB medium as a reference, alternatively.

The proteomic response of *E. coli* was analyzed with respect to the associated biological function revealing remarkable differences in the regulation of transcription and translation. Most notably, a generic response typically linked to temperature-induced and osmotic stress was observed upon deuteration. Among these response genes, chaperons acting on RNA and DNA were identified, which suggests an alteration in the stability of nucleic acids caused by incorporation of deuterium. Significant differences in the expression of the ribosome-related control genes such as *raiA* and *rpsD* suggest increased occurrence of misreading events during translation. This observation further supports the idea of a deuterium-related change in nucleic acid stability. Moreover, several chaperons involved in protein folding displayed higher expression levels in a perdeuterated environment. The hydrophobic effect of the solvent is possibly enlarged in D_2O (*196*) and may increase the aggregation propensity as

well as the number of non-native conformations sampled during protein folding. Adaptation to growth in a perdeuterated environment may be further influenced by abiotic factors other than the properties of the solvent such as the aeration and the pH of the culture. To identify such an influence, tight control of these parameters is required, but only limited in culture flasks. Fermenter-controlled growth of *E. coli* in perdeuterated minimal medium may enable to identify a possible proteomic response to changes in the aeration and the pH on the adaptation.

Furthermore, alteration of the cellular energy metabolism was observed in D_2O-adapted cells, which is in agreement with the recently reported changes in the TCA (*175*). The proteome-wide analysis also revealed changes in the oxidative phosphorylation, which was not accessible to the previous metabolic analysis by ^{13}C-NMR. Whereas a significant change of the aerobic energy generation was indicated by the differences in the expression of proteins in the electron transport chain, enzymes required for gluconeogenesis and the glyoxylate pathway were simultaneously upregulated. This metabolic switch is likely caused by a kinetic isotope effect, which may affect rate-limiting steps in glycolysis and the TCA. The overall energy demand of D_2O-adapted cells may increase due to elevated chaperone activity and error correction mechanisms related to improper translation.

Eventually, the approach may guide the development of optimized protocols for deuterated expression in *E. coli*, which are aimed at improving the protein yield by genetic modification to increase resistance to D_2O and refinement of deuterated minimal media. Particularly, supplementation of acetate is likely to overcome deficiencies in the energy metabolism solely based on glucose (*184*). Alternatively, growth on an increased amount of glucose may also be possible to improve the protein yield. In addition, genetic modification may be performed by long-term adaptation to a perdeuterated minimal medium to select for random mutations, which promote growth and expression under this condition. Further, it is hoped that proteomic analysis can be employed to identify the limitations of deuterated expression in higher eukaryotes. Common to eukaryotic cells is their high degree of internal organization in functional compartments such as the nucleus and mitochondria, which are bounded by intracellular membranes (*38*). As inferred from the proteomic response of *E. coli*, osmotic stress in a perdeuterated environment is likely to change the size of these compartments and may interfere with their function. The preliminary results obtained for *E. coli* also indicate a central role of anaplerotic reactions to handle the increased demand of energy, which was apparent from differential expression of proteins involved in the electron transport chain. Interestingly, the existence of a glyoxylate shunt has not been established for insect and mammalian cells (*224*). In theory, the required enzymes malate synthase and isocitrate lyase may be transiently introduced during expression using the baculovirus (*78*). A major concern with respect to the observed increase in the level of chaperones is a possible cell cycle arrest of higher eukaryotic cells due to significant alteration of DNA and RNA stability, which is likely to ultimately inhibit growth in a perdeuterated medium. Hence, like for *E. coli*, stepwise adaptation to high levels of D_2O is expected to be required for the higher eukaryotic expression systems.

A possible extension of the described approach may complement the proteomic analysis with changes in the transcriptome and even allow for detection of alternative splicing patterns and gene fusion (*225*). Indeed, changes in the mRNA level may not correlate to the observed difference in protein expression (*226*) and thus might reveal additional insights into gene regulation. Moreover, stability-related changes in the protein abundance may be identified

with this combined approach as deuteration is reported to improve or worsen the stability of proteins based on the nature of the weak interaction (*189*, *196*).

Material and Methods

Growth and adaptation of E. coli

Cultures of *E. coli* BL21(D3) cells were grown on a rotary shaker (220 rpm, 50 mm throw) at 37 °C in batches of 10 ml in 100 ml Erlenmeyer flasks to provide sufficient aeration. Adaptation to growth in D_2O was achieved by repeated subculturing. First, cultures grown to stationary phase in LB medium were used to inoculate M9 minimal medium at an OD_{600} of 0.1. Subsequently, cells in protonated M9 medium were grown for about 5 h to the end of the exponential growth phase and samples were taken for whole-cell lysis. To reduce the spillover of unlabeled medium, cells were separated from the supernatant by centrifugation at 5000*g* for 5 min. The pellet was carefully resuspended in M9 medium containing 70 % D_2O to yield an OD_{600} of 0.1 and cultures were grown for about 7.5 h to the end of the exponential growth phase. Finally, cells adapted to growth in 70 % D_2O were transferred to 99.8 % D_2O M9 medium containing 2 g/L $^2H_7$-glucose (98 %-d_7, Sigma-Aldrich) as the sole carbon source (M9-D). Again, cultures were inoculated to yield an OD_{600} of 0.1 and centrifugation was used to limit dilution of the isotope label. M9-D cultures were grown for about 12 h to the end of the exponential growth phase and samples were taken for whole-cell lysis. Samples of M9 and M9-D cultures were normalized for the determined OD_{600}.

Whole-cell lysis and tryptic digestion of E. coli

Cell pellets were resuspended in 100 mM ammonium carbonate containing 2 % sodium deoxycholate (SDC) and 5 mM TCEP (lysis buffer). The cells were lysed by combined harsh vortexing (3 × 30 s) and sonication (3 × 10 s, 0.5 cycle and 100 % amplitude) using a VialTweeter (Hielscher). Obtained lysates were heated at 95 °C for 10 min shaken at 500 rpm, followed by centrifugation at 10.000*g* for 10 s to remove cell debris. The protein concentration in the supernatant was determined using a BCA assay kit (Thermo Fisher Scientific). Subsequently, samples were diluted in lysis buffer to yield protein concentration of 5 µg/µl and alkylation was performed on 100 µg of total protein in a volume of 20 µl using 10 mM of freshly prepared iodoacetamide for 30 min at 25 °C in the dark. The excess of iodoacetamide was quenched by addition of 12 mM *N*-acetyl-cysteine.

For digestion, samples were diluted to yield a final SDC concentration of 1 % using 100 mM ammonium carbonate. Subsequently, trypsin (Sigma-Aldrich) was added in a 1:50 enzyme/protein ratio digestion was performed over night at 37 °C. The reaction was terminated by addition of 1 % trifluoroacetic acid, followed by centrifugation to remove SDC precipitate at 15.000*g* for 10 min (4 °C). Centrifugation was repeated once more and the obtained supernatant was used for solid phase extraction.

Peptides were desalted on a C18 reversed-phase spin column according to the supplier's manual (Macrospin, Harvard Apparatus). Three biological replicates of cells grown in M9 and M9-D were normalized by their peptide content and combined for LC/MS analysis.

LC-MS/MS analysis

Each sample was subjected to LC–MS analysis using a dual pressure LTQ-Orbitrap Elite mass spectrometer connected to an electrospray ion source (Thermo Fisher Scientific) as described recently (*102*) with a few modifications. In brief, peptide separation was carried out using an EASY nLC-1000 system (Thermo Fisher Scientific) equipped with a RP-HPLC column (75 µm × 45 cm) packed in-house with C18 resin (ReproSil-Pur C18–AQ, 1.9 µm resin; Dr. Maisch GmbH, Ammerbuch-Entringen, Germany) using a linear gradient from 95 % solvent A (0.15 % formic acid, 2% acetonitrile) and 5 % solvent B (98 % acetonitrile, 0.15 % formic acid) to 28 % solvent B over 120 min at a flow rate of 0.2 µl/min. The data acquisition mode was set to obtain one high resolution MS scan in the FT part of the mass spectrometer at a resolution of 240,000 full width at half-maximum (at m/z 400) followed by MS/MS scans in the linear ion trap of the 20 most intense ions. The charged state screening modus was enabled to exclude unassigned and singly charged ions and the dynamic exclusion duration was set to 30 s. The ion accumulation time was set to 300 ms (MS) and 50 ms (MS/MS).

Identification and quantification of non-deuterated peptides

MS raw files were imported into the *Progenesis QI* LC-MS software (Nonlinear Dynamics, Version v.2.0) and analyzed using the default parameter settings. MS/MS data were exported directly from *Progenesis* in mgf format and searched using *Mascot* (Matrix Science, Version 2.4.0) against a concatenated target-decoy *E. coli* protein database including forward and reversed sequences (UniProt, release date 01.01.2015 including 4479 target sequences). The search criteria were set as follows: 10 ppm precursor ion mass tolerance, 0.6 Da fragment ion mass tolerance, and full tryptic specificity required (cleavage after lysine or arginine residues); maximum 3 missed cleavages; fixed modification: carbamidomethylation (C), variable modification: oxidation (M). The database search results were filtered limiting the peptide and protein level false discovery rate (FDR) to 1 %. The *Mascot* peptide identifications were subsequently imported into *Progenesis* and mapped to the detected MS1 peaks to ≤90 %.

Relative quantification of deuterated and non-deuterated proteins

Peak identification and quantification for deuterated peptides was obtained using an in-house algorithm for exact detection of deuterated isotopic envelopes (EDDIE) written in *Python* 2.7. Theoretical isotope patterns of deuterated peptides were calculated as binomials accounting for residual protonation and matched with the observed m/z values in the MS1 data. The retention times of the light peptides determined by *Progenesis QI* were used to define a retention time window for the corresponding deuterated peptides. Relative quantification was based on the intensities of the monoisotopic mass of the light peptide and the five most abundant deuterated isotopomers excluding natural abundance contributions.

Peptide annotated MS1 peak intensity lists were processed to obtain protein level relative abundances using the *SafeQuant R package* (available on CRAN) (*102*). Global data normalization was achieved by equalizing the total MS1 peak intensities across all LC-MS runs and channels. Subsequently, normalized MS1 peak intensities were combined per protein and LC-MS/MS run to calculate the protein abundance ratios followed by testing for

differential expression using a moderated t-statistic $\tilde{t}$ based on empirical Bayes variance estimators (*182*) according to:

$$\tilde{t} = \frac{1}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \frac{\hat{\beta}}{\tilde{s}_g} \quad [1]$$

where n_i denotes the number of replicates in the *i*th group and $\hat{\beta}$ represents the least-square estimator of the $\log_2$ fold change in expression for a particular gene. The posterior gene-specific (*g*) standard deviation $\tilde{s}_g$ is calculated using:

$$\tilde{s}_g = \sqrt{\frac{d_0 {s_0}^2 + d_g {s_g}^2}{d_0 + d_g}} \quad [2]$$

where d_g and ${s_g}^2$ correspond to the statistical degrees of freedom and variance of the experiment, and d_0 and ${s_0}^2$ are the prior degrees of freedom and variance estimated from a chi-square distribution, respectively. The associated probability values (p-values) were calculated from the moderated t-statistic $\tilde{t}$ (Equation 1) based on the t-distribution using $d_0 + d_g$ degrees of freedom (*227*). These p-values, which reflect the probability of detecting a given mean expression difference across the sample conditions by chance alone were corrected for multiple testing to obtain the false discovery rates (q-values) according to the Benjamini-Hochberg method (*183*). The final candidate list of differentially expressed proteins was obtained using a 2-fold abundance ratio cut-off and a q-value cut-off of 0.01 reflecting a false-discovery rate of ≤1 %.

Functional annotation and classification

Identification of functionally related gene groups in the final candidate list were obtained using the *DAVID* bioinformatics database (v6.7) (*185*). Gene functional annotation was analyzed using default medium and high classification stringency settings.

8 Economic production of isotope-labeled proteins in mammalian cells using cellular extracts of yeast and algae

Introduction

Recent progress in gene expression and cultivation protocols of mammalian cell lines (MCL) has raised their popularity for structural studies of complex proteins (*3*, *228*). Recombinant expression in mammalian cells has been prominently applied to produce biologically active G protein coupled receptors for crystallographic and NMR studies due to the native cellular environment and a comprehensive post-translational modification machinery(*229*). MCL have been improved to match the scalability of insect cell suspension cultures and are amenable to growth in synthetically defined serum- and protein-free expression media. Furthermore, transient transfection of MCL dispenses with the need for laborious generation of recombinant baculovirus, which is generally required for baculovirus-infected insect cells.

In theory, similar considerations as discussed for isotope labeling in baculovirus-infected insect cells apply to MCL (Chapter 1). Most importantly, the use of complex, chemically defined growth media is required due to limitations of the amino acid metabolism in animal cells. Customized dropout media are available from commercial suppliers but can as well be prepared in-house based on inexpensive reagents (*230*), albeit at the expense of additional labor. Major difficulties of the insect cell system also apply to MCL, namely high costs of practically all labeling approaches and the absence of an efficient protocol for uniform deuteration.

Amino acid type selective labeling has been successfully applied to NMR studies of isotope-labeled bovine rhodopsin expressed in HEK293 cells (*231-234*). Labeling by multiple combinations of ^{13}C and ^{15}N enabled resonance assignment of all five tryptophan residues by solid-state NMR (*234*). In addition, simultaneous labeling of several amino acids (Gln, Gly, Leu, Lys, Ser, Thr, Trp, and Val) was achieved by the use of commercial growth media (*235*).

Uniform isotope labeling in MCL was pioneered in CHO and Sp2/0 cells by Hansen and co-workers more than 20 years ago (*32*). The described approach enabled labeling by ^{15}N and ^{13}C based on the combination of bacterial and algal cell extracts. These cells can be grown on inexpensive minimal media containing $^{15}NH_4Cl$ and ^{13}C-glucose or $^{13}CO_2$, respectively. However, the extracts were produced by acid hydrolysis and further purification was essential to promote growth of the MCL. Moreover, acid hydrolysis results in degradation of amino acids such as glutamine and cysteine, which in turn had to be supplied in the mammalian growth medium.

Today, purified algal amino acid mixtures are readily available from commercial suppliers at moderate costs. These mixtures have been successfully applied for NMR studies but still required additional modifications of the medium, most importantly supplementation of arginine, tryptophan and tyrosine (*236*, *237*). Efficient isotope incorporation (>90 %) has been relying on at least two subsequent passages of the MCL in labeling medium prior to expression of the target protein.

In general, replacement of all amino acid sources and their respective precursors such as carbohydrates in the medium by the individual labeled components is a straightforward approach to uniform isotope labeling. Growth media for MCL entirely based on supplementing labeled amino acid sources have been commercialized and provide uniform ^{15}N and ^{13}C,^{15}N labeling at costs of 7700 EUR and 15400 EUR per liter, respectively (BioExpress-6000, Cambridge Isotope Laboratories). Although these media are available for several years and have been successfully used to produce isotope-labeled bovine rhodopsin in

HEK293 cells (*235*), they have so far not found routine application in part due to high costs and the absence of deuterated labeling schemes.

A recent development to provide an alternative medium for uniform isotope labeling has been based on a combination of yeast and algal extracts (*238*). This approach has enabled uniform labeling in HEK293 cells by the use of a modified DMEM/F12 medium formulation. However, four full passages in the labeled medium have been required to achieve equally high incorporation (~90 %) as reported for previous protocols. Furthermore, labeling has only been limited to ^{15}N and application to canonical NMR experimental schemes has not been demonstrated.

This chapter describes the initial application of isotope-labeled yeast extracts for economic production of uniformly labeled proteins in mammalian cells. Recombinant expression of the mouse angiotensin II receptor type 1a (mAT_1) was analyzed in HEK293 cells grown on commercially available dropout media supplemented with extracts of yeast and algae as the single source of amino acids. The method enabled high-level expression of mAT_1 corresponding to about 60 % of the yield obtained in commercial media by direct exchange to extract-based growth media prior to expression. Finally, possible optimizations of the protocol by adaptation of HEK293 cells to the extract-based medium are discussed.

Remark on author contribution

The work of this chapter was carried out together with Dr. Chayne Piscitelli (Paul Scherrer Institute, Villigen), who performed cloning and expression of the mouse angiotensin II receptor type 1a in HEK293S cells.

Results and Discussion

Initial design of the extract-based growth media

To establish the amount of amino acids required as a supplement to an amino acid-depleted medium for HEK293S cells, concentrations of individual amino acids in the widely used Dulbeccos's modified Eagle's minimal essential medium (DMEM) were used as a reference. When compared to the overall content of amino acids in insect cell media such as IPL-41 (Chapter 2.1), DMEM provides a significantly lower amount of amino acids (Figure 8.1, grey bars). Furthermore, A, D, E, N, and P are not present in DMEM. These amino acids are non-essential and thus can be synthesized by mammalian cells. In contrast, high levels of glutamine (584 mg/L) are added to DMEM, which is in line with the supplementation scheme developed for insect cells grown on yeast extracts (Chapter 2.1).

Initially, the use of an amino acid-depleted dropout medium derived from the commercial protein expression medium (PEM, Thermo Fisher) was considered as this medium was successfully employed to obtain high-level expression of the angiotensin II receptor type 1a (mAT_1). Unfortunately, customized formulations of PEM are costly (450 EUR/L, personal communication with Thermo Fisher). Therefore, an inexpensive dropout medium depleted of all amino acids (ΔM77, Bioconcept) based on the commercial MAM-PF 77 medium (M77, Bioconcept) was chosen as an alternative to PEM. As the amino acid content of M77 is not disclosed, their content was determined by amino acid analysis as described in Chapter 2. Using the same assumptions as made for dropout media supplemented with yeast extracts used in insect cells, the amino acid content of M77 was analyzed at 20 % (w/v) of the initial amino acid level (Figure 8.1, brown bars). Remarkably, the analysis indicated that in such a medium most amino acids would exceed the concentrations of the DMEM reference. Moreover, M77 did not contain alanine and glutamine. The latter is commonly added shortly before use due to the instability of the amino acid (*138*), thereby increasing the shelf life of these media. The concentration of cysteine was not determined by this method (see Chapter 2.1) and a possible limitation cannot be excluded. However, supplementation of cysteine was not required in extract-based growth media for insect cells, which contain an overall higher amount of amino acids.

In a next step, the amount of yeast extract necessary for supplementation to the amino acid-depleted ΔM77 was determined. Comparison to DMEM suggested that in most cases supplementation of ΔM77 with 8 g/L yeast extract will provide the required concentrations of amino acids (Figure 8.1, blue bars). Since the concentrations of isoleucine and tyrosine were only slightly lower, an influence on the growth of HEK293S cells was not expected in such a medium. Conversely, the concentration of glutamine was significantly lower (~60-fold) in the yeast extract and would have required addition of substantially higher amounts of the extract. Therefore, ΔM77 supplemented with 8 g/L yeast extract and 1 g/L glutamine was taken as an initial medium (ΔM77+YE) for expression of mAT_1 in HEK293S cells.

In order to test for possible growth promoting (*142*, *143*) or inhibitory effects of the yeast extract, an alternative supplementation scheme was developed using algal extract (AE) as described for insect cells in Chapter 4. Since the amino acid analysis of AE indicated that the overall content of amino acids was similar to yeast extract (see Chapter 4), the same supplementation scheme (8 g/L AE and 1 g/L glutamine) was used for preparation of ΔM77+AE. The algal extract-based medium also required additional supplementation of 20 mg/L tryptophan (*39*).

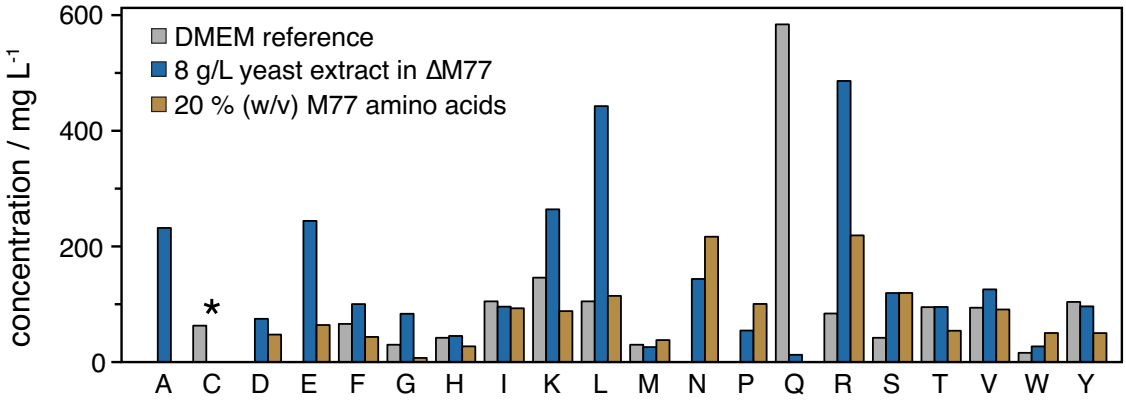


Figure 8.1 Amino acid content of media used for growth of HEK293S cells in suspension. The free amino acid content of the commercial DMEM (*grey*) is compared to 20 % concentrations of amino acids in commercial MAM-PF 77 (M77, *brown*) and a customized amino acid-depleted M77 medium (ΔM77) supplemented with 8 g/L yeast extract (*blue*). Concentrations of amino acids in DMEM were taken from (*239*). Free amino acids in homemade yeast extract and M77 were quantified by amino acid analysis (see Chapter 2.1). Levels of cysteine were not determined by this method (asterisk).

Expression of mAT$_1$ in HEK293S cells using cellular extracts of yeast and algae

To test the performance of amino acid-depleted ΔM77 supplemented with cellular extracts of yeast or algae, a stabilized construct of the mouse angiotensin II receptor type 1a (TS-mAT$_1$) was expressed in HEK293S cells grown on different expression media. TS-mAT$_1$ comprises a C-terminal fusion protein eGFP (see Materials and Methods), which enables to analyze the yield and the oligomerization state of the receptor using fluorescence-detection size-exclusion chromatography (FSEC). Typically, ~0.3 mg/L of purified TS-mAT$_1$ were obtained by expression in the commercial protein expression medium (PEM, Thermo Fisher).

Expression of TS-mAT$_1$ in ΔM77 requires medium exchange from PEM prior to induction, which may affect cell viability and the protein yield. To test for such effects, HEK293S cells grown to a cell density of 1.5×10^6 cells ml^{-1} in PEM were collected by a mild centrifugation step (see Materials and Methods) and again resuspended in fresh PEM. FSEC analysis of solubilized TS-mAT$_1$ expressed in these cells (Figure 8.2, PEM) indicated a single peak for the monomeric receptor at an elution volume of about 2.55 ml (Figure 8.2, grey rectangle). Formation of higher oligomers could also be excluded from this analysis, as the corresponding peak was not observed at an elution volume of 1.9 ml. Furthermore, a second peak was observed at an elution volume of about 3.3 ml, which most likely corresponds to free eGFP. Substantial amounts of eGFP were so far not observed for expression in PEM (data not shown) and suggest an alteration of the expression due to the exchange of the medium. However, it is expected that free eGFP is removed during purification. The expression level obtained by medium exchange in PEM was further used as a reference to estimate the amount of TS-mAT$_1$ expressed in HEK293S cells grown on extract-based media.

In a next step, cultures of HEK293S cells were first grown in PEM and exchanged to ΔM77 supplemented with 8 g/L yeast extract (YE) and 1 g/L glutamine. Again, expression was analyzed on the solubilized receptor using FSEC (Figure 8.2, ΔM77+YE). The yield of monomeric TS-mAT$_1$ was reduced to about 50 % compared to the reference expression level in PEM. Interestingly, a clear doublet was shown for the peak possibly corresponding to free

eGFP. This observation may be related to growth in ΔM77. Moreover, a decrease in the overall fluorescence in the FSEC measurement indicates a reduction of the total yield of protein.

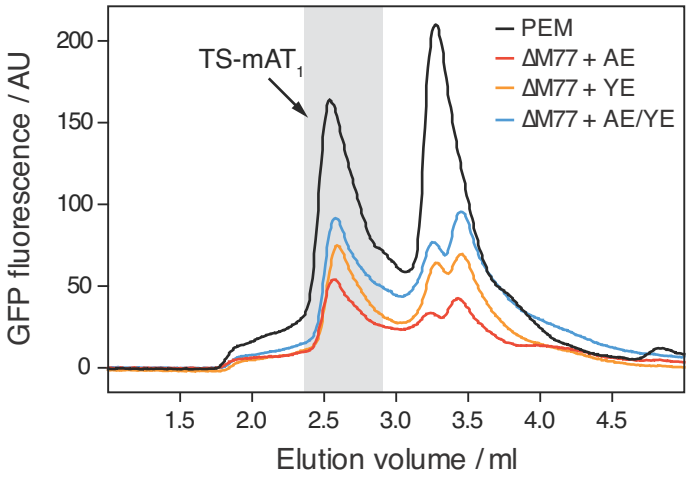


Figure 8.2 Fluorescence-detection size-exclusion chromatography (FSEC) analysis of a stabilized mouse angiotensin II receptor type 1a (TS-mAT$_1$) solubilized in *n*-dodecyl β-D-maltopyranoside. TS-mAT$_1$ comprises a C-terminal fusion to the green fluorescence protein (eGFP) to enable fluoresence detection of the receptor (see Materials and Methods). The solubilized receptor was obtained from HEK293S cells grown on the commercial protein expression medium (PEM, Thermo Fisher) and the amino acid-depleted MAM-PF 77 medium (ΔM77, Bioconcept) supplemented with 8 g/L yeast extract (ΔM77+YE), 8 g/L algal extract (ΔM77+AE), and a combination of 4 g/L of each extract (ΔM77+AE/YE).

A similar FSEC profile was obtained for TS-mAT$_1$ expressed HEK293S cells grown in ΔM77 supplemented with 8 g/L algal extract (AE), 1 g/L glutamine, and 20 mg/L tryptophan (Figure 8.2, ΔM77+AE). Compared to expression in PEM, the yield of monomeric receptor was significantly reduced to about 30 %. Again, the overall fluorescence intensity was reduced and a clear doublet was observed for the eGFP peak. Since amino acid analysis of indicated similar amounts of free amino acids in AE and YE, the reduction may not be simply explained by a limitation of single amino acids. AE contains substantial levels of free ammonium (*39*), which is known to be toxic to mammalian cell cultures at high concentrations (*240*). Therefore, decreasing the amount of AE supplemented to ΔM77 may improve expression in this medium. Furthermore, as discussed above, supplementation of YE to growth media used for CHO cells significantly improved proliferation and protein expression (*142*, *143*). A similar growth-promoting effect is likely to be expected in HEK293S cultures.

To analyze the expression level for a combination of both extracts, TS-mAT$_1$ was expressed in HEK293S cells grown in ΔM77 supplemented with 1 g/L glutamine and 4 g/L YE and AE, respectively. FSEC analysis indicated that the peak of the monomeric receptor was reduced to about 60 % relative to expression in PEM (Figure 8.2, ΔM77+AE/YE). Accordingly, this corresponds to 2-fold increase in the yield over ΔM77+AE. The improvement may be attributed to the growth-promoting property of YE. However, as the amount of AE was simultaneously decreased, a possible inhibitory effect of the AE cannot be ruled out from the data. Although expression of TS-mAT$_1$ was slightly higher in

ΔM77+AE/YE compared to ΔM77+YE, the difference may be explained by a naturally occurring variation of the expression level. Interestingly, a similar decrease in the protein yield has been reported in a previous protocol using AE and YE (*238*), which hints at a more fundamental limitation related to the use of such extracts. However, robust expression in HEK293S and CHO cells has not clearly been demonstrated for this method in part due to batch-to-batch variations of the prepared extracts.

Furthermore, glutamine is a major source of ammonium, which is released to the culture medium during growth and is toxic to mammalian cells (*240*). According to the amino acid content of DMEM (Figure 8.1, grey bars), a reduction of the initial concentration of glutamine from 1 g/L to 584 mg/L may be possible without affecting cell growth in ΔM77. Consequently, the amount of secreted ammonium may be reduced and expression levels may be further improved.

Adaptation of HEK293 cells to growth in alternative media

For all tested supplementation schemes based on ΔM77, expression of TS-mAT$_1$ was significantly reduced in HEK293S cells. To investigate whether such a decrease might be caused by the medium alone, adaptation of HEK293S cells to growth in commercial MAM-PF 77 (M77) was investigated. First, cell growth was monitored for HEK293S cells cultured in PEM supplemented with 5 % FCS (Figure 8.3A, PEM+FCS). A cell density of 1.5×10^6 cells ml^{-1} was reached seven days post inoculation, which is in good agreement with typically obtained growth rates in PEM+FCS. In contrast, M77 supplemented with 5 % FCS did not support growth of HEK293S cells (Figure 8.3A, M77+FCS). Remarkably, a complete growth arrest was observed after three days of culturing. The cells appeared shriveled and smaller in size compared to the cultures grown in PEM+FCS. This hints on osmotic stress, which is typically caused by a change in the osmolarity of the growth medium. M77 is optimized for serum-free growth of CHO cells (personal communication with R. Liebetanz, Bioconcept). Supplementation of FCS is likely to increase the osmolarity of the medium, which may exceed the level tolerated by HEK293S. To test for such an effect of FCS, HEK293S cells were grown in M77 without supplementation of FCS. This serum-free medium supported growth of the culture up to a final cell density of 0.72×10^6 cells ml^{-1} (Figure 8.3A, M77). A significant reduction in the growth rate was observed after three days of culturing. This effect was more pronounced from day five on indicating an apparent stationary behavior. Such a reduction of the growth rate at a later stage of the culture may be explained by two reasons: (i) a depletion of a single or several nutrients and (ii) accumulation of an inhibitory byproduct such as ammonium. A possible limitation of amino acids may be identified by amino acid analysis of M77 at this time point. Furthermore, the cell size was reduced by about 2-fold compared to cells grown in PEM+FCS. Again, this hints on a difference in the osmolarity between M77 and PEM, which should be tested in a next step.

To further study the effect on the protein yield, expression of TS-mAT$_1$ was induced in M77 and PEM+FCS cultures seven days post inoculation (Figure 8.3A, green rectangle). The detergent-soluble supernatant obtained from HEK293S cells after 72 h of expression was separated by SDS-PAGE and TS-mAT$_1$ was detected by in-gel fluorescence (see Materials and Methods). Analysis of the Coomassie-stained gel indicated a significant reduction of the overall expression in the supernatant of HEK293S cells grown on M77 compared to cultures in PEM+FCS (Figure 8.3B, SDS-PAGE). Furthermore, a strong fluorescent band

corresponding to the monomeric TS-mAT$_1$ was detected in the supernatant of detergent-solubilized HEK293S cells grown on PEM+FCS using in-gel fluorescence (Figure 8.3B, iGF). Conversely, fluorescence of eGFP was hardly detectable in the supernatant obtained from M77 cultures (Figure 8.3B, M77). Spectrometric quantification of the eGFP fluorescence in the two supernatants revealed that the expression of TS-mAT$_1$ was reduced to 6 % in HEK293S cells grown on M77 as compared to PEM+FCS cultures.

In summary, the data suggest that M77 is not optimized for growth of HEK293S cells and causes a decrease in the protein yield irrespective of supplementation with cellular extracts of yeast or algae.

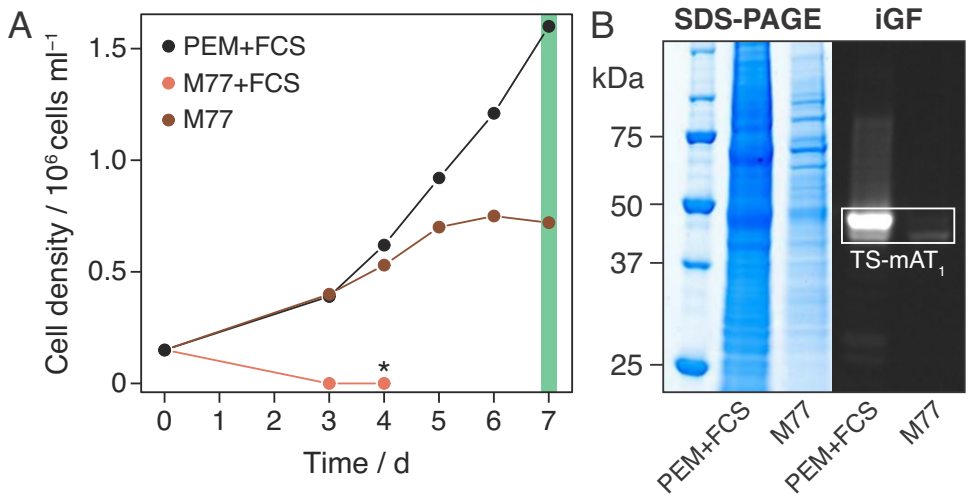


Figure 8.3 A Growth of HEK293S cells in different commercial media. Cultures inoculated with 0.15×10^6 cells ml^{-1} were grown on protein expression medium (PEM, Thermo Fisher) supplemented with 5 % FCS (PEM+FCS), MAM-PF 77 (M77, Bioconcept) supplemented with 5 % FCS (M77+FCS), and M77 alone. The M77+FCS culture was no longer monitored after day four as growth was arrested as indicated by the asterisk. Expression of TS-mAT$_1$ was induced seven days post inoculation (green rectangle). **B** Analysis of TS-mAT$_1$ expression in HEK293S cells grown on PEM+FCS and M77. Detergent-soluble supernatants were seperated by SDS-PAGE and TS-mAT$_1$ was detected by in-gel fluorescence (iGF) analysis of the C-terminal fusion protein eGFP (see Materials and Methods).

Conclusion and perspective

Recombinant expression of the stabilized mouse angiotensin II receptor type 1a (TS-mAT$_1$) was established in HEK293S cells using cellular extracts of yeast or algae supplemented to inexpensive dropout media depleted of amino acids. The method was adapted from the protocol originally described for production of isotope-labeled proteins in insect cells (Chapter 2.1) and thus may enable uniform isotope labeling of TS-mAT$_1$ in an economic manner. So far, the highest yield was achieved by the combined use of yeast extract (YE) and algal extract (AE) added to ΔM77 corresponding to 60 % of the expression obtained in the full medium. However, also supplementation of YE alone resulted in comparable expression of TS-mAT$_1$ (~50 %). Surprisingly, amino acid analysis of the YE did not indicate significant limitations of individual amino acids, which have been observed for extract-based labeling media used for recombinant expression in insect cells (Chapter 2.1). It

is expected that the method may be improved by modifying the supplementation scheme based on the amino acid analysis of the cellular extracts.

Furthermore, adaptation of HEK293S cells to growth in the commercial MAM-PF 77 medium (M77) used as the basis for the extract-based dropout media was studied. As shown here, expression of TS-mAT$_1$ was dramatically decreased in this medium. Since adaptation to growth in the commercial M77 was so far not successful, different media formulations need to be explored. The preparation of a homemade growth medium provides a promising alternative to commercially available media. Such homemade media have already been described for NMR studies of G protein coupled receptors (*239*). Amino acid analysis of the commercial protein expression medium (PEM) may guide optimization of the amino acid content in these homemade media.

In the next step, isotope-labeled extracts can be employed to produce uniformly labeled TS-mAT$_1$ in HEK293S cells. A thorough analysis of the receptor by mass spectrometry is required to characterize isotope incorporation (Chapter 2). Mass spectrometric analysis is possibly impeded for the fusion construct of TS-mAT$_1$ due to the requirement of detergent solubilization. However, isotope incorporation may readily be determined on the green fluorescent protein (eGFP) fused to the C-terminus of the receptor, which can be obtained in a soluble form by thrombin cleavage of TS-mAT$_1$. This may enable quantification of the overall as well as the amino acid resolved isotope incorporation.

It is also expected that the use of deuterated YE or AE may provide labeling by ^{2}H in HEK293S cells and eventually leads to similar incorporation as described for the labeling protocol in insect cells (Chapter 2.1). This may be demonstrated by NMR analysis of ^{2}H,^{15}N-labeled TS-mAT$_1$ expressed in HEK293S cells grown on ^{2}H,^{15}N-labeled YE. As shown in Chapter 3.3, deuteration has been essential to obtain high-quality NMR spectra of the ^{2}H,^{15}N-labeled β_1-adrenergic receptor, which has been solubilized in detergent micelles and hence resembles a system of similar complexity.

Finally, it is hoped that the described approach may be developed into a robust method for uniform isotope labeling in mammalian cell lines including CHO cells. In the case of the CHO cell line, the use of the commercial MAM-PF 77 medium as a basis for the extract-based labeling media may be beneficial, since this medium was originally developed for recombinant expression in CHO cells.

Material and Methods

Construct design of the mouse angiotensin II receptor

A stabilized construct of the mouse angiotensin II receptor type 1a (TS-mAT$_1$) was derived from the wild-type receptor sequence (*agtr1a*) by addition of the single point mutation F117W to enhance expression and obtain a putative increase in thermostability. As compared to the wild-type receptor, TS-mAT$_1$ further contains a truncation at the C-terminus (Δ324). The modified receptor sequence was fused at the C-terminus via a thrombin cleavage site to the eGFP reporter gene sequence (*241*) and the 1D4 epitope tag for purification (*242*, *243*) to yield the final construct TS-mAT$_1$-eGFP/1D4. This construct was cloned into the pACMV-*TetO* plasmid to enable expression under the control of the CMV promoter based on a tetracycline-inducible operator (*TetO*) sequence (*244*).

Stable transfection and growth of HEK293S cells

Generation and selection of HEK293S GnTI cells stably transfected with TS-mAT$_1$-eGFP/1D4 was carried out as described previously (*245*). Stable cell lines were expanded in 15-cm-diameter culture dishes at 37 °C using 5 % CO_2 in DMEM (Thermo Fisher) containing 10 % FCS, 10 µg/ml blasticidin, and 1 mg/ml G418. Adherent cells were treated with trypsin at a confluency of 80 % and subsequently washed from the culture dish using PEM (Thermo Fisher) supplemented with 5 % FCS and 1x GlutaMAX™ (Thermo Fisher). Suspension cultures were inoculated at a cell density of 0.15×10^6 cells ml^{-1} and grown at 37 °C in Erlenmeyer flasks shaken at 110 rpm using 5 % CO_2.

Quantitative amino acid analysis by LC/MS

Detailed protocols are presented in Chapter 2.1.

Growth media containing extracts of yeast and algal cells

Dropout media were based on the commercial MAM-PF 77 medium (Bioconcept) depleted of all amino acids (ΔM77). Supplementation of ΔM77 with cellular extracts of algae and yeast was adapted from the protocols used in insect cells, which are presented in Chapter 4 and Chapter 2.1. For optimal growth of HEK293S cells in ΔM77, pH and osmolarity were adjusted to 7.4 and 300 mosmol kg^{-1}, respectively.

Expression of TS-mAT$_1$-eGFP/1D4 in dropout media supplemented with cellular extracts

HEK293S GnTI stably transfected with TS-mAT$_1$-eGFP/1D4 were seeded in PEM supplemented with 5 % FCS and 1x GlutaMAX™ and cultured to a cell density of 1.5×10^6 cells ml^{-1}. Subsequently, cells were collected by centrifugation at 300*g* for 5 min and resuspended in 50 ml of ΔM77 supplemented with different cellular extracts. Expression was induced by addition of 2 µg/ml tetracycline and 5 mM sodium butyrate. Moreover, 1 µM of the antagonist azilsartan was added to improve the yield of the receptor. The cells were further incubated for 48 h and 1 ml of culture was removed to test for expression.

Analysis of TS-mAT$_1$-eGFP/1D4 expression

Cells collected after 48 h of expression were harvested by centrifugation at 3000*g* for 30 s (4 °C) and resuspended in 200 µl of solubilization buffer containing 50 mM HEPES, 300 mM NaCl, 20 mM *n*-dodecyl β-D-maltopyranoside, 2 mM CHS, 100 µM azilsartan and a protease inhibitor cocktail (complete, Roche). Solubilization was performed at 4 °C for 1.5 h using mild agitation, followed by an ultracentrifugation step at 90000*g* for 30 min (4 °C) to remove insoluble material. For fluorescence-detection size-exclusion chromatography (FSEC) analysis, 10 µl of the supernatant was injected onto a Tosoh SuperSW3000 column (4.6 x 300 mm) using an Äkta Ettan LC system (GE Healthcare) running at 0.2 mL/min and 4°C in SEC buffer containing 50 mM HEPES pH 7.0, 300 mM NaCl, 1 mM DDM, 0.1 mM CHS, and 10 µM azilsartan. The elution profile of TS-mAT$_1$-eGFP/1D4 was monitored using an in-line fluorimeter to detect GFP fluorescence at 520 nm. For in-gel fluorescence analysis, 10 µl of the supernatant was separated by SDS-PAGE on a 12 % TEA-Tricine SDS PAGE gel under reducing conditions. Subsequently, the gel was exposed for 1 min using an

Amersham Imager 600 (GE Healthcare) to detect fluorescence of eGFP. The total protein was visualized using the InstantBlue™ stain (Sigma-Aldrich).

Adaptation of HEK293S cells to growth in commercial dropout media

HEK293S GnTI stably transfected with TS-mAT$_1$-eGFP/1D4 were first expanded in DMEM as described above and subsequently seeded at a cell density of 0.15×10^6 cells ml^{-1} in MAM-PF 77 (Bioconcept) supplemented with 1x GlutaMAX™ and either with or without 5 % FCS. Suspension cultures were grown for seven days at 37 °C in Erlenmeyer flasks shaken at 110 rpm using 5 % CO_2.

References

1. L. E. Kay, K. H. Gardner, Solution NMR spectroscopy beyond 25 kDa. *Curr Opin Struct Biol*. **7**, 722–731 (1997).

2. K. H. Gardner, L. E. Kay, The use of 2H, 13C, 15N multidimensional NMR to study the structure and dynamics of proteins. *Annu Rev Biophys Biomol Struct*. **27**, 357–406 (1998).

3. R. Assenberg, P. T. Wan, S. Geisse, L. M. Mayr, Advances in recombinant protein expression for use in pharmaceutical research. *Curr Opin Struct Biol*. **23**, 393–402 (2013).

4. A. de Marco, Strategies for successful recombinant expression of disulfide bond-dependent proteins in Escherichia coli. *Microb Cell Fact*. **8**, 26–26 (2008).

5. O. Kolaj, S. Spada, S. Robin, J. G. Wall, Use of folding modulators to improve heterologous protein production in Escherichia coli. *Microb Cell Fact*. **8**, 9 (2009).

6. M. Wacker, D. Linton, P. G. Hitchen, M. Nita-Lazar, S. M. Haslam, S. J. North, M. Panico, H. R. Morris, A. Dell, B. W. Wren, M. Aebi, N-linked glycosylation in Campylobacter jejuni and its functional transfer into E. coli. *Science*. **298**, 1790–1793 (2002).

7. V. D. Nguyen, F. Hatahet, K. E. H. Salo, E. Enlund, C. Zhang, L. W. Ruddock, Pre-expression of a sulfhydryl oxidase significantly increases the yields of eukaryotic disulfide bond containing proteins expressed in the cytoplasm of E.coli. *Microb Cell Fact*. **10**, 1 (2011).

8. T. A. Kost, J. P. Condreay, D. L. Jarvis, Baculovirus as versatile vectors for protein expression in insect and mammalian cells. *Nat. Biotechnol*. **23**, 567–575 (2005).

9. D. Massotte, G protein-coupled receptor overexpression with the baculovirus-insect cell system: a tool for structural and functional studies. *Biochim. Biophys. Acta*. **1610**, 77–89 (2003).

10. T. Saarenpää, V.-P. Jaakola, A. Goldman, Baculovirus-mediated expression of GPCRs in insect cells. *Meth. Enzymol*. **556**, 185–218 (2015).

11. W. Weber, E. Weber, S. Geisse, K. Memmert, Optimisation of protein expression and establishment of the Wave Bioreactor for Baculovirus/insect cell culture. *Cytotechnology*. **38**, 77–85 (2001).

12. V. Singh, Disposable bioreactor for cell culture using wave-induced agitation. *Cytotechnology*. **30**, 149–158 (1999).

13. A. Strauss, F. Bitsch, G. Fendrich, P. Graff, R. Knecht, B. Meyhack, W. Jahnke, Efficient uniform isotope labeling of Abl kinase expressed in Baculovirus-infected insect cells. *J. Biomol. NMR*. **31**, 343–349 (2005).

14. K. Saxena, A. Dutta, J. Klein-Seetharaman, H. Schwalbe, Isotope labeling in insect cells. *Methods Mol. Biol.* **831**, 37–54 (2011).

15. A. D. Gossert, W. Jahnke, Isotope labeling in insect cells. *Adv. Exp. Med. Biol.* **992**, 179–196 (2012).

16. T. A. Egorova-Zachernyuk, G. J. C. G. M. Bosman, A. M. A. Pistorius, W. J. DeGrip, Production of yeastolates for uniform stable isotope labelling in eukaryotic cell culture. *Appl Microbiol Biotechnol.* **84**, 575–581 (2009).

17. T. A. Egorova-Zachernyuk, G. J. C. G. M. Bosman, W. J. DeGrip, V. I. Shvets, Stable isotope labelling of human histamine receptor H1R: prospects for structure-based drug design. *Dokl. Biochem. Biophys.* **433**, 164–167 (2010).

18. A. Strauss, F. Bitsch, B. Cutting, G. Fendrich, P. Graff, J. Liebetanz, M. Zurini, W. Jahnke, Amino-acid-type selective isotope labeling of proteins expressed in Baculovirus-infected insect cells useful for NMR studies. *J. Biomol. NMR.* **26**, 367–372 (2003).

19. A. D. Gossert, A. Hinniger, S. Gutmann, W. Jahnke, A. Strauss, C. Fernández, A simple protocol for amino acid type selective isotope labeling in insect cells with improved yields and high reproducibility. *J. Biomol. NMR.* **51**, 449–456 (2011).

20. L. Skora, B. Shrestha, A. D. Gossert, Isotope Labeling of Proteins in Insect Cells. *Meth. Enzymol.* **565**, 245–288 (2015).

21. S. A. Weiss, G. C. Smith, S. S. Kalter, J. L. Vaughn, Improved method for the production of insect cell cultures in large volume. *In Vitro.* **17**, 495–502 (1981).

22. A. F. Creemers, C. H. Klaassen, P. H. Bovee-Geurts, R. Kelle, U. Kragl, J. Raap, W. J. de Grip, J. Lugtenburg, H. J. de Groot, Solid state 15N NMR evidence for a complex Schiff base counterion in the visual G-protein-coupled receptor rhodopsin. *Biochemistry.* **38**, 7195–7199 (1999).

23. M. Brüggert, T. Rehm, S. Shanker, J. Georgescu, T. A. Holak, A novel medium for expression of proteins selectively labeled with 15N-amino acids in Spodoptera frugiperda (Sf9) insect cells. *J. Biomol. NMR.* **25**, 335–348 (2003).

24. A. Strauss, G. Fendrich, M. A. Horisberger, J. Liebetanz, B. Meyhack, J.-M. Schlaeppi, R. Schmitz, Improved expression of kinases in Baculovirus-infected insect cells upon addition of specific kinase inhibitors to the culture helpful for structural studies. *Protein Expr Purif.* **56**, 167–176 (2007).

25. N. Vajpai, A. Strauss, G. Fendrich, S. W. Cowan-Jacob, P. W. Manley, W. Jahnke, S. Grzesiek, Backbone NMR resonance assignment of the Abelson kinase domain in complex with imatinib. *Biomol NMR Assign.* **2**, 41–42 (2008).

26. N. Vajpai, A. Strauss, G. Fendrich, S. W. Cowan-Jacob, P. W. Manley, S. Grzesiek, W. Jahnke, Solution conformations and dynamics of ABL kinase-inhibitor complexes determined by NMR substantiate the different binding modes of imatinib/nilotinib and dasatinib. *J. Biol. Chem.* **283**, 18292–18302 (2008).

27. M. Kainosho, T. Tsuji, Assignment of the three methionyl carbonyl carbon resonances in Streptomyces subtilisin inhibitor by a carbon-13 and nitrogen-15 double-labeling technique. A new strategy for structural studies of proteins in solution. *Biochemistry*. **21**, 6273–6279 (1982).

28. A. Bax, F. Delaglio, S. Grzesiek, G. W. Vuister, Resonance assignment of methionine methyl groups and chi 3 angular information from long-range proton-carbon and carbon-carbon J correlation in a calmodulin-peptide complex. *J. Biomol. NMR*. **4**, 787–797 (1994).

29. Y. Kofuku, T. Ueda, J. Okude, Y. Shiraishi, K. Kondo, M. Maeda, H. Tsujishita, I. Shimada, Efficacy of the β_2-adrenergic receptor is determined by conformational equilibrium in the transmembrane region. *Nat Commun*. **3**, 1045–1045 (2011).

30. L. Ohman, J. Ljunggren, L. Haggstrom, Induction of a metabolic switch in insect cells by substrate-limited fed batch cultures. *Appl Microbiol Biotechnol*. **43**, 1006–1013 (1995).

31. J. Hamatsu, D. O'Donovan, T. Tanaka, T. Shirai, Y. Hourai, T. Mikawa, T. Ikeya, M. Mishima, W. Boucher, B. O. Smith, E. D. Laue, M. Shirakawa, Y. Ito, High-resolution heteronuclear multidimensional NMR of proteins in living insect cells using a baculovirus protein expression system. *J. Am. Chem. Soc*. **135**, 1688–1691 (2013).

32. A. P. Hansen, A. M. Petros, A. P. Mazar, T. M. Pederson, A. Rueter, S. W. Fesik, A practical method for uniform isotopic labeling of recombinant proteins in mammalian cells. *Biochemistry*. **31**, 12713–12718 (1992).

33. C. F. Shen, T. Kiyota, B. Jardin, Y. Konishi, A. Kamen, Characterization of yeastolate fractions that promote insect cell growth and recombinant protein production. *Cytotechnology*. **54**, 25–34 (2007).

34. D. M. LeMaster, F. M. Richards, Preparative-scale isolation of isotopically labeled amino acids. *Anal. Biochem*. **122**, 238–247 (1982).

35. M. Drews, M. Doverskog, L. Ohman, B. E. Chapman, U. Jacobsson, P. W. Kuchel, L. Haggstrom, Pathways of glutamine metabolism in Spodoptera frugiperda (Sf9) insect cells: evidence for the presence of the nitrogen assimilation system, and a metabolic switch by 1H/15N NMR. *J. Biotechnol*. **78**, 23–37 (2000).

36. M. Doverskog, L. Han, L. Haggstrom, Cystine/cysteine metabolism in cultured Sf9 cells: influence of cell physiology on biosynthesis, amino acid uptake and growth. *Cytotechnology*. **26**, 91–102 (1998).

37. L. Öhman, M. Alarcon, J. Ljunggren, A.-K. Ramqvist, L. Häggström, Glutamine is not an essential amino acid for Sf-9 insect cells. *Biotechnol. Lett*. **18**, 765–770 (1996).

38. D. Voet, J. G. Voet, *Biochemistry, 4th Edition* (John Wiley & Sons, 2010).

39. A. Sitarska, L. Skora, J. Klopp, S. Roest, C. Fernández, B. Shrestha, A. D. Gossert,

Affordable uniform isotope labeling with (2)H, (13)C and (15)N in insect cells. *J. Biomol. NMR*. **62**, 191–197 (2015).

40. A. Meola, C. Deville, S. A. Jeffers, P. Guardado-Calvo, I. Vasiliauskaite, C. Sizun, C. Girard-Blanc, C. Malosse, C. van Heijenoort, J. Chamot-Rooke, T. Krey, E. Guittet, S. Pêtres, F. A. Rey, F. Bontems, Robust and low cost uniform 15N-labeling of proteins expressed in Drosophila S2 cells and Spodoptera frugiperda Sf9 cells for NMR applications. *J Struct Biol*. **188**, 71–78 (2014).

41. Y. Kofuku, T. Ueda, J. Okude, Y. Shiraishi, K. Kondo, T. Mizumura, S. Suzuki, I. Shimada, Functional Dynamics of Deuterated β2‐Adrenergic Receptor in Lipid Bilayers Revealed by NMR Spectroscopy. *Angewandte Chemie*. **126**, 13594–13597 (2014).

42. T. J. Kurtti, S. P. Chaudhary, M. A. Brooks, Influence of physical factors on the growth of insect cells in vitro. I. Effect of osmotic pressure on growth rate of a moth cell line. *In Vitro*. **10**, 149–156 (1974).

43. E. J. Schlaeger, Medium design for insect cell culture. *Cytotechnology*. **20**, 57–70 (1996).

44. J. L. Vaughn, R. H. Goodwin, G. J. Tompkins, P. McCawley, The establishment of two cell lines from the insectspodoptera frugiperda (lepidoptera; noctuidae). *In Vitro*. **13**, 213–217 (1977).

45. I. Schneider, Cell lines derived from late embryonic stages of Drosophila melanogaster. *Development*. **27**, 353–365 (1972).

46. Y. T. Chung, E. B. Keller, Positive and negative regulatory elements mediating transcription from the Drosophila melanogaster actin 5C distal promoter. *Mol. Cell. Biol*. **10**, 6172–6180 (1990).

47. G. Maroni, E. Otto, D. Lastowski-Perry, Molecular and cytogenetic characterization of a metallothionein gene of drosophila. *Genetics*. **112**, 493–504 (1986).

48. H. Johansen, A. van der Straten, R. Sweet, E. Otto, G. Maroni, M. Rosenberg, Regulated expression at high copy number allows production of a growth-inhibitory oncogene product in Drosophila Schneider cells. *Genes Dev*. **3**, 882–889 (1989).

49. A. M. Moraes, S. A. C. Jorge, R. M. Astray, C. A. T. Suazo, C. E. Calderón Riquelme, E. F. P. Augusto, A. Tonso, M. M. Pamboukian, R. A. M. Piccoli, M. F. Barral, C. A. Pereira, Drosophila melanogaster S2 cells for expression of heterologous genes: From gene cloning to bioprocess development. *Biotechnol. Adv*. **30**, 613–628 (2012).

50. L. A. Guarino, M. D. Summers, Nucleotide sequence and temporal expression of a baculovirus regulatory gene. *J. Virol*. **61**, 2091–2099 (1987).

51. D. D. Hegedus, T. A. Pfeifer, J. Hendry, D. A. Theilmann, T. A. Grigliatti, A series of broad host range shuttle vectors for constitutive and inducible expression of heterologous proteins in insect cell lines. *Gene*. **207**, 241–249 (1998).

52. V. A. Morais, J. Costa, Stable expression of recombinant human alpha3/4 fucosyltransferase III in Spodoptera frugiperda Sf9 cells. *J. Biotechnol.* **106**, 69–75 (2003).

53. S. Turan, C. Zehe, J. Kuehle, J. Qiao, J. Bode, Recombinase-mediated cassette exchange (RMCE) - a rapidly-expanding toolbox for targeted genomic modifications. *Gene.* **515**, 1–27 (2013).

54. M. Kito, S. Itami, Y. Fukano, K. Yamana, T. Shibui, Construction of engineered CHO strains for high-level production of recombinant proteins. *Appl Microbiol Biotechnol.* **60**, 442–448 (2002).

55. F. Fernandes, J. Vidigal, M. M. Dias, K. L. J. Prather, A. S. Coroadinha, A. P. Teixeira, P. M. Alves, Flipase-mediated cassette exchange in Sf9 insect cells for stable gene expression. *Biotechnol. Bioeng.* **109**, 2836–2844 (2012).

56. G. E. Smith, M. J. Fraser, M. D. Summers, Molecular Engineering of the Autographa californica Nuclear Polyhedrosis Virus Genome: Deletion Mutations Within the Polyhedrin Gene. *J. Virol.* **46**, 584–593 (1983).

57. G. E. Smith, M. D. Summers, M. J. Fraser, Production of human beta interferon in insect cells infected with a baculovirus expression vector. *Mol. Cell. Biol.* **3**, 2156–2165 (1983).

58. M. M. van Oers, G. P. Pijlman, J. M. Vlak, Thirty years of baculovirus-insect cell protein expression: from dark horse to mainstream technology. *J. Gen. Virol.* **96**, 6–23 (2015).

59. R. R. Granados, L. Guoxun, A. C. G. Derksen, K. A. McKenna, A new insect cell line from Trichoplusia ni (BTI-Tn-5B1-4) susceptible to Trichoplusia ni single enveloped nuclear polyhedrosis virus. *J Invertebr Pathol.* **64**, 260–266 (1994).

60. M. B. Keith, P. J. Farrell, K. Iatrou, L. A. Behie, Screening of transformed insect cell lines for recombinant protein production. *Biotechnol. Prog.* **15**, 1046–1052 (1999).

61. R. A. Taticek, C. Choi, S. E. Phan, L. A. Palomares, M. L. Shuler, Comparison of growth and recombinant protein expression in two different insect cell lines in attached and suspension culture. *Biotechnol. Prog.* **17**, 676–684 (2001).

62. T. J. Wickham, T. Davis, R. R. Granados, M. L. Shuler, H. A. Wood, Screening of insect cell lines for the production of recombinant proteins and infectious virus in the baculovirus expression system. *Biotechnol. Prog.* **8**, 391–396 (1992).

63. T. Kato, M. Kajikawa, K. Maenaka, E. Y. Park, Silkworm expression system as a platform technology in life science. *Appl Microbiol Biotechnol.* **85**, 459–470 (2010).

64. H. Yagi, M. Nakamura, J. Yokoyama, Y. Zhang, T. Yamaguchi, S. Kondo, J. Kobayashi, T. Kato, E. Y. Park, S. Nakazawa, N. Hashii, N. Kawasaki, K. Kato, Stable isotope labeling of glycoprotein expressed in silkworms using immunoglobulin G as a test molecule. *J. Biomol. NMR.* **62**, 157–167 (2015).

65. P. V. Vail, G. Sutter, D. L. Jay, D. Gough, Reciprocal infectivity of nuclear

polyhedrosis viruses of the cabbage looper and alfalfa looper. *J Invertebr Pathol*. **17**, 383–388 (1971).

66. J. F. Henderson, P. Faulkner, E. A. MacKinnon, Some biophysical properties of virus present in tissue cultures infected with the nuclear polyhedrosis virus of Trichoplusia ni. *J. Gen. Virol*. **22**, 143–146 (1974).

67. L. E. Volkman, M. D. Summers, C. H. Hsieh, Occluded and nonoccluded nuclear polyhedrosis virus grown in Trichoplusia ni: comparative neutralization comparative infectivity, and in vitro growth studies. *J. Virol*. **19**, 820–832 (1976).

68. B. G. Ooi, L. K. Miller, Regulation of Host Rna Levels During Baculovirus Infection. *Virology*. **166**, 515–523 (1988).

69. D. L. Knudson, K. A. Harrap, Replication of nuclear polyhedrosis virus in a continuous cell culture of Spodoptera frugiperda: microscopy study of the sequence of events of the virus infection. *J. Virol*. **17**, 254–268 (1975).

70. E. B. Carstens, S. T. Tjia, W. Doerfler, Infection of Spodoptera frugiperda cells with Autographa californica nuclear polyhedrosis virus I. Synthesis of intracellular proteins after virus infection. *Virology*. **99**, 386–398 (1979).

71. S. Ghosh, A. Jain, B. Mukherjee, S. Habib, S. E. Hasnain, The host factor polyhedrin promoter binding protein (PPBP) is involved in transcription from the baculovirus polyhedrin gene promoter. *J. Virol*. **72**, 7484–7493 (1998).

72. J. Kuzio, D. Z. Rohel, C. J. Curry, A. Krebs, E. B. Carstens, P. Faulkner, Nucleotide sequence of the p10 polypeptide gene of Autographa californica nuclear polyhedrosis virus. *Virology*. **139**, 414–418 (1984).

73. G. E. Smith, J. M. Vlak, M. D. Summers, Physical Analysis of Autographa-Californica Nuclear Polyhedrosis-Virus Transcripts for Polyhedrin and 10,000-Molecular-Weight Protein. *J. Virol*. **45**, 215–225 (1983).

74. M. M. VanOers, J. M. Vlak, The baculovirus 10-kDa protein. *J Invertebr Pathol*. **70**, 1–17 (1997).

75. J. P. Burand, M. D. Summers, G. E. Smith, Transfection with baculovirus DNA. *Virology*. **101**, 286–290 (1980).

76. V. A. Luckow, S. C. Lee, G. F. Barry, P. O. Olins, Efficient generation of infectious recombinant baculoviruses by site-specific transposon-mediated insertion of foreign genes into a baculovirus genome propagated in Escherichia coli. *J. Virol*. **67**, 4566–4579 (1993).

77. P. A. Kitts, R. D. Possee, A method for producing recombinant baculovirus expression vectors at high frequency. *BioTechniques*. **14**, 810–817 (1993).

78. I. Berger, D. J. Fitzgerald, T. J. Richmond, Baculovirus expression system for heterologous multiprotein complexes. *Nat. Biotechnol*. **22**, 1583–1587 (2004).

79. Y. Kanai, T. N. Athmaram, M. Stewart, P. Roy, Multiple large foreign protein

expression by a single recombinant baculovirus: a system for production of multivalent vaccines. *Protein Expr Purif*. **91**, 77–84 (2013).

80. D. J. Wasilko, S. E. Lee, K. J. Stutzman-Engwall, B. A. Reitz, T. L. Emmons, K. J. Mathis, M. J. Bienkowski, A. G. Tomasselli, H. D. Fischer, The titerless infected-cells preservation and scale-up (TIPS) method for large-scale production of NO-sensitive human soluble guanylate cyclase (sGC) from insect cells infected with recombinant baculovirus. *Protein Expr Purif*. **65**, 122–132 (2009).

81. M. M. van Oers, Opportunities and challenges for the baculovirus expression system. *J Invertebr Pathol*. **107 Suppl**, S3–15 (2011).

82. R. E. Hawtin, T. Zarkowska, K. Arnold, C. J. Thomas, G. W. Gooday, L. A. King, J. A. Kuzio, R. D. Possee, Liquefaction of Autographa californica nucleopolyhedrovirus-infected insects is dependent on the integrity of virus-encoded chitinase and cathepsin genes. *Virology*. **238**, 243–253 (1997).

83. C. J. Thomas, H. L. Brown, C. R. Hawes, B. Y. Lee, M. K. Min, L. A. King, R. D. Possee, Localization of a baculovirus-induced chitinase in the insect cell endoplasmic reticulum. *J. Virol*. **72**, 10207–10212 (1998).

84. R. D. Possee, C. J. Thomas, L. A. King, The use of baculovirus vectors for the production of membrane proteins in insect cells. *Biochem Soc Trans*. **27**, 928–932 (1999).

85. J. M. Slack, J. Kuzio, P. Faulkner, Characterization of v-cath, a cathepsin L-like proteinase expressed by the baculovirus Autographa californica multiple nuclear polyhedrosis virus. *J. Gen. Virol*. **76 (Pt 5)**, 1091–1098 (1995).

86. S. A. Kaba, A. M. Salcedo, P. O. Wafula, J. M. Vlak, M. M. van Oers, Development of a chitinase and v-cathepsin negative bacmid for improved integrity of secreted recombinant proteins. *J Virol Methods*. **122**, 113–118 (2003).

87. L. G. Hom, L. E. Volkman, Autographa californica M nucleopolyhedrovirus chiA is required for processing of V-CATH. *Virology*. **277**, 178–183 (2000).

88. M. J. Betenbaugh, N. Tomiya, S. Narang, J. T. Hsu, Y. C. Lee, Biosynthesis of human-type N-glycans in heterologous systems. *Curr Opin Struct Biol*. **14**, 601–606 (2004).

89. L. Kong, N. C. Sheppard, G. B. E. Stewart-Jones, C. L. Robson, H. Chen, X. Xu, G. Krashias, C. Bonomelli, C. N. Scanlan, P. D. Kwong, S. A. Jeffs, I. M. Jones, Q. J. Sattentau, Expression-system-dependent modulation of HIV-1 envelope glycoprotein antigenicity and immunogenicity. *J. Mol. Biol*. **403**, 131–147 (2010).

90. A. Helenius, M. Aebi, Roles of N-linked glycans in the endoplasmic reticulum. *Annu Rev Biochem*. **73**, 1019–1049 (2004).

91. K. Breitbach, D. L. Jarvis, Improved glycosylation of a foreign protein by Tn-5B1-4 cells engineered to express mammalian glycosyltransferases. *Biotechnol. Bioeng*. **74**, 230–239 (2001).

92. J. R. Hollister, D. L. Jarvis, Engineering lepidopteran insect cells for sialoglycoprotein production by genetic transformation with mammalian beta 1,4-galactosyltransferase and alpha 2,6-sialyltransferase genes. *Glycobiology*. **11**, 1–9 (2001).

93. Y. Zhao, D. A. G. Chapman, I. M. Jones, Improving baculovirus recombination. *Nucleic Acids Res*. **31**, E6–6 (2003).

94. J. Mitsuhashi, *Invertebrate Tissue Culture Methods* (Springer Science & Business Media, 2012).

95. M. Doverskog, U. Jacobsson, B. E. Chapman, P. W. Kuchel, L. Haggstrom, Determination of NADH-dependent glutamate synthase (GOGAT) in Spodoptera frugiperda (Sf9) insect cells by a selective 1H/15N NMR in vitro assay. *J. Biotechnol*. **79**, 87–97 (2000).

96. V. Bernal, N. Carinhas, A. Y. Yokomizo, M. J. T. Carrondo, P. M. Alves, Cell density effect in the baculovirus-insect cells system: A quantitative analysis of energetic metabolism. *Biotechnol. Bioeng*. **104**, 162–180 (2009).

97. D. T. Wong, R. W. Fuller, B. B. Molloy, Inhibition of amino acid transaminases by L-cycloserine. *Adv. Enzyme Regul*. **11**, 139–154 (1973).

98. C. Opitz, S. Isogai, S. Grzesiek, An economic approach to efficient isotope labeling in insect cells using homemade 15N-, 13C- and 2H-labeled yeast extracts. *J. Biomol. NMR*. **62**, 373–385 (2015).

99. M. K. Rosen, K. H. Gardner, R. C. Willis, W. E. Parris, T. Pawson, L. E. Kay, Selective methyl group protonation of perdeuterated proteins. *J. Mol. Biol*. **263**, 627–636 (1996).

100. T. Etezady-Esfarjani, S. Hiller, C. Villalba, K. Wüthrich, Cell-free protein synthesis of perdeuterated proteins for NMR studies. *J. Biomol. NMR*. **39**, 229–238 (2007).

101. J. Yokoyama, T. Matsuda, S. Koshiba, N. Tochio, T. Kigawa, A practical method for cell-free protein synthesis to avoid stable isotope scrambling and dilution. *Anal. Biochem*. **411**, 223–229 (2011).

102. T. Glatter, C. Ludwig, E. Ahrné, R. Aebersold, A. J. R. Heck, A. Schmidt, Large-Scale Quantitative Assessment of Different In-Solution Protein Digestion Protocols Reveals Superior Cleavage Efficiency of Tandem Lys-C/Trypsin Proteolysis over Trypsin Digestion. *J. Proteome Res*. **11**, 5145–5156 (2012).

103. S. Isogai, X. Deupi, C. Opitz, F. M. Heydenreich, C.-J. Tsai, F. Brueckner, G. F. X. Schertler, D. B. Veprintsev, S. Grzesiek, Backbone NMR reveals allosteric signal transduction networks in the β1-adrenergic receptor. *Nature*. **530**, 237–241 (2016).

104. E. Ghosh, P. Kumari, D. Jaiman, A. K. Shukla, Methodological advances: the unsung heroes of the GPCR structural revolution. *Nat Rev Mol Cell Biol*. **16**, 69–81 (2015).

105. T. Warne, R. Moukhametzianov, J. G. Baker, R. Nehmé, P. C. Edwards, A. G. W. Leslie, G. F. X. Schertler, C. G. Tate, The structural basis for agonist and partial

agonist action on a β1-adrenergic receptor. *Nature*. **469**, 241–244 (2011).

106. T. Warne, M. J. Serrano-Vega, J. G. Baker, R. Moukhametzianov, P. C. Edwards, R. Henderson, A. G. W. Leslie, C. G. Tate, G. F. X. Schertler, Structure of a β1-adrenergic G-protein-coupled receptor. *Nature*. **454**, 486–491 (2008).

107. M. P. Bokoch, Y. Zou, S. G. F. Rasmussen, C. W. Liu, R. Nygaard, D. M. Rosenbaum, J. J. Fung, H.-J. Choi, F. S. Thian, T. S. Kobilka, J. D. Puglisi, W. I. Weis, L. Pardo, R. S. Prosser, L. Mueller, B. K. Kobilka, Ligand-specific regulation of the extracellular surface of a G-protein-coupled receptor. *Nature*. **463**, 108–112 (2010).

108. S. H. Park, B. B. Das, F. Casagrande, Y. Tian, H. J. Nothnagel, M. Chu, H. Kiefer, K. Maier, A. A. De Angelis, F. M. Marassi, S. J. Opella, Structure of the chemokine receptor CXCR1 in phospholipid bilayers. *Nature*. **491**, 779–783 (2012).

109. R. Nygaard, Y. Zou, R. O. Dror, T. J. Mildorf, D. H. Arlow, A. Manglik, A. C. Pan, C. W. Liu, J. J. Fung, M. P. Bokoch, F. S. Thian, T. S. Kobilka, D. E. Shaw, L. Mueller, R. S. Prosser, B. K. Kobilka, The dynamic process of β(2)-adrenergic receptor activation. *Cell*. **152**, 532–542 (2013).

110. J. L. Miller, C. G. Tate, Engineering an ultra-thermostable β(1)-adrenoceptor. *J. Mol. Biol*. **413**, 628–638 (2011).

111. M. J. Serrano-Vega, F. Magnani, Y. Shibata, C. G. Tate, Conformational thermostabilization of the beta1-adrenergic receptor in a detergent-resistant form. *Proc. Natl. Acad. Sci. U.S.A*. **105**, 877–882 (2008).

112. F. Brueckner, C. L. Piscitelli, C.-J. Tsai, J. Standfuss, X. Deupi, G. F. X. Schertler, Structure of β-adrenergic receptors. *Meth. Enzymol*. **520**, 117–151 (2013).

113. A. Rath, M. Glibowicka, V. G. Nadeau, G. Chen, C. M. Deber, Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. *Proc. Natl. Acad. Sci. U.S.A*. **106**, 1760–1765 (2009).

114. A. I. Alexandrov, M. Mileni, E. Y. T. Chien, M. A. Hanson, R. C. Stevens, Microscale fluorescent thermal stability assay for membrane proteins. *Structure*. **16**, 351–359 (2008).

115. F. C. Ayers, G. L. Warner, K. L. Smith, D. A. Lawrence, Fluorometric quantitation of cellular and nonprotein thiols. *Anal. Biochem*. **154**, 186–193 (1986).

116. J. A. Goncalves, K. South, S. Ahuja, E. Zaitseva, C. A. Opefi, M. Eilers, R. Vogel, P. J. Reeves, S. O. Smith, Highly conserved tyrosine stabilizes the active state of rhodopsin. *Proc. Natl. Acad. Sci. U.S.A*. **107**, 19861–19866 (2010).

117. X. Deupi, J. Standfuss, Structural insights into agonist-induced activation of G-protein-coupled receptors. *Curr Opin Struct Biol*. **21**, 541–551 (2011).

118. D. R. O'Reilly, L. K. Miller, V. A. Luckow, *Baculovirus expression vectors: a laboratory manual* (Oxford University Press, 1994).

119. M. Ikura, L. E. Kay, A. Bax, A novel approach for sequential assignment of 1H, 13C, and 15N spectra of proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. *Biochemistry*. **29**, 4659–4667 (1990).

120. S. Grzesiek, A. Bax, Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. *Journal of Magnetic Resonance (1969)*. **96**, 432–440 (1991).

121. M. Salzmann, K. Pervushin, G. Wider, H. Senn, K. Wüthrich, TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 13585–13590 (1998).

122. D. Marion, P. C. Driscoll, L. E. Kay, P. T. Wingfield, A. Bax, A. M. Gronenborn, G. M. Clore, Overcoming the overlap problem in the assignment of 1H NMR spectra of larger proteins by use of three-dimensional heteronuclear 1H-15N Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1 beta. *Biochemistry*. **28**, 6150–6156 (1989).

123. D. Marion, L. E. Kay, S. W. Sparks, D. A. Torchia, A. Bax, Three-dimensional heteronuclear NMR of nitrogen-15 labeled proteins. *J. Am. Chem. Soc.* **111**, 1515–1517 (1989).

124. K. V. Pervushin, G. Wider, R. Riek, K. Wüthrich, The 3D NOESY-[(1)H,(15)N,(1)H]-ZQ-TROSY NMR experiment with diagonal peak suppression. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 9607–9612 (1999).

125. K. Wüthrich, *NMR of Proteins and Nucleic Acids* (Wiley-Interscience, 1986).

126. D. Nietlispach, R. T. Clowes, R. W. Broadhurst, Y. Ito, J. Keeler, M. Kelly, J. Ashurst, H. Oschkinat, P. J. Domaille, E. D. Laue, An Approach to the Structure Determination of Larger Proteins Using Triple Resonance NMR Experiments in Conjunction with Random Fractional Deuteration. *J. Am. Chem. Soc.* **118**, 407–415 (1996).

127. X. Ding, X. Zhao, A. Watts, G-protein-coupled receptor structure, ligand binding and activation as studied by solid-state NMR spectroscopy. *Biochem. J.* **450**, 443–457 (2013).

128. F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, A. Bax, NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*. **6**, 277–293 (1995).

129. T. D. Goddard, G. D. Kneller, SPARKY 3 (2008).

130. E.-Y. Yun, T.-W. Goo, S.-W. Kim, K.-H. Choi, J.-S. Hwang, S.-W. Kang, O.-Y. Kwon, Changes in cellular secretory processing during baculovirus infection. *Biotechnol. Lett.* **27**, 1041–1045 (2005).

131. C.-Y. Teng, S.-L. Chang, M. M. van Oers, T.-Y. Wu, Enhanced protein secretion from insect cells by co-expression of the chaperone calreticulin and translation initiation factor eIF4E. *Mol. Biotechnol.* **54**, 68–78 (2013).

132. J. K. Eckert, Y. J. Kim, J. I. Kim, K. Gürtler, D.-Y. Oh, S. Sur, L. Lundvall, L.

Hamann, A. van der Ploeg, P. Pickkers, E. Giamarellos-Bourboulis, A. V. Kubarenko, A. N. Weber, M. Kabesch, O. Kumpf, H.-J. An, J.-O. Lee, R. R. Schumann, The crystal structure of lipopolysaccharide binding protein reveals the location of a frequent mutation that impairs innate immunity. *Immunity*. **39**, 647–660 (2013).

133. D. L. Jarvis, C. Oker-Blom, M. D. Summers, Role of glycosylation in the transport of recombinant glycoproteins through the secretory pathway of lepidopteran insect cells. *J. Cell. Biochem.* **42**, 181–191 (1990).

134. D. L. J. Xianzong Shi, Protein N-Glycosylation in the Baculovirus-Insect Cell System. *Current drug targets*. **8**, 1116 (2007).

135. M. Fountoulakis, H. W. Lahm, Hydrolysis and amino acid composition of proteins. *J Chromatogr A*. **826**, 109–134 (1998).

136. J. M. Brown, Compositions and methods for protein structural determinations (1997).

137. C. B. Airaudo, A. Gayte-Sorbier, P. Armand, Stability of Glutamine and Pyroglutamic Acid under Model System Conditions: Influence of Physical and Technological Factors. *J Food Science*. **52**, 1750–1752 (1987).

138. G. L. Tritsch, G. E. Moore, Spontaneous decomposition of glutamine in cell culture media. *Exp. Cell Res.* **28**, 360–364 (1962).

139. M. K. Snowden, J. H. Baxter, M. Mamula Bergana, I. Reyzer, V. Pound, Stability of N-Acetylglutamine and Glutamine in Aqueous Solution and in a Liquid Nutritional Product by an Improved HPLC Method. *J Food Science*. **67**, 384–389 (2002).

140. B. J. Smith, *Protein Sequencing Protocols* (Springer Science & Business Media, 2002).

141. J. F. Sissom, L. Ellis, Biosynthesis of the precursor of a soluble human insulin receptor ectodomain in insect Sf9 cells infected with a recombinant baculovirus. *Biochem. Biophys. Res. Commun.* **177**, 764–770 (1991).

142. M. Mosser, R. Kapel, A. Aymes, L.-M. Bonanno, E. Olmos, I. Chevalot, I. Marc, A. Marc, Chromatographic fractionation of yeast extract: A strategy to identify physicochemical properties of compounds promoting CHO cell culture. *Process Biochem.* **47**, 1178–1185 (2012).

143. M. Mosser, R. Kapel, I. Chevalot, E. Olmos, I. Marc, A. Marc, E. Oriol, Fractionation of yeast extract by nanofiltration process to assess key compounds involved in CHO cell culture improvement. *Biotechnol Progress*. **31**, 875–882 (2015).

144. M. Mosser, I. Chevalot, E. Olmos, F. Blanchard, R. Kapel, E. Oriol, I. Marc, A. Marc, Combination of yeast hydrolysates to improve CHO cell growth and IgG production. *Cytotechnology*. **65**, 629–641 (2013).

145. V. Sklenář, A. Bax, Spin-echo water suppression for the generation of pure-phase

two-dimensional NMR spectra. *Journal of Magnetic Resonance (1969)*. **74**, 469–479 (1987).

146. C. D. Listrom, H. Morizono, B. S. Rajagopal, M. T. McCann, M. Tuchman, N. M. Allewell, Expression, purification, and characterization of recombinant human glutamine synthetase. *Biochem. J.* **328 (Pt 1)**, 159–163 (1997).

147. J. Neermann, R. Wagner, Comparative analysis of glucose and glutamine metabolism in transformed mammalian cell lines, insect and primary liver cells. *J. Cell. Physiol.* **166**, 152–169 (1996).

148. D. M. LeMaster, F. M. Richards, NMR sequential assignment of Escherichia coli thioredoxin utilizing random fractional deuteriation. *Biochemistry* (1988).

149. D. M. LeMaster, Deuterium labelling in NMR structural analysis of larger proteins. *Q Rev Biophys.* **23**, 133–174 (1990).

150. V. Kragl, A. Gödde, C. Wandrey, W. Kinzy, J. J. Cappon, J. Lugtenburg, Repetitive batch as an efficient method for preparative scale enzymic synthesis of 5-azido-neuraminic acid and 15N-l-glutamic acid. *Tetrahedron: Asymmetry*. **4**, 1193–1202 (1993).

151. C. A. Woolfolk, B. Shapiro, E. R. Stadtman, Regulation of glutamine synthetase. I. Purification and properties of glutamine synthetase from Escherichia coli. *Arch. Biochem. Biophys.* **116**, 177–192 (1966).

152. P. C. Engel, K. Dalziel, Kinetic studies of glutamate dehydrogenase. The reductive amination of 2-oxoglutarate. *Biochem. J.* **118**, 409–419 (1970).

153. A. Radzicka, R. Wolfenden, Comparing the polarities of the amino acids: side-chain distribution coefficients between the vapor phase, cyclohexane, 1-octanol, and neutral aqueous solution. *Biochemistry*. **27**, 1664–1670 (1988).

154. J. Singh, M. C. Joshi, R. Bhatnagar, Cloning and expression of mycobacterial glutamine synthetase gene in Escherichia coli. *Biochem. Biophys. Res. Commun.* **317**, 634–638 (2004).

155. D. T. Browne, G. L. Kenyon, E. L. Packer, H. Sternlicht, D. M. Wilson, Studies of macromolecular structure by 13 C nuclear magnetic resonance. II. A specific labeling approach to the study of histidine residues in proteins. *J. Am. Chem. Soc.* **95**, 1316–1323 (1973).

156. T. Quinto, D. Häussinger, V. Köhler, T. R. Ward, Artificial metalloenzymes for the diastereoselective reduction of NAD +to NAD 2H. *Org. Biomol. Chem.* **13**, 357–360 (2015).

157. R. Mehta, J. T. Pearson, S. Mahajan, A. Nath, M. J. Hickey, D. R. Sherman, W. M. Atkins, Adenylylation and catalytic properties of Mycobacterium tuberculosis glutamine synthetase expressed in Escherichia coli versus mycobacteria. *J. Biol. Chem.* **279**, 22477–22482 (2004).

158. J. D. Gawronski, D. R. Benson, Microtiter assay for glutamine synthetase

biosynthetic activity using inorganic phosphate detection. *Anal. Biochem.* **327**, 114–118 (2004).

159. T. D. Meek, J. J. Villafranca, Kinetic mechanism of Escherichia coli glutamine synthetase. *Biochemistry*. **19**, 5513–5519 (1980).

160. G. Harth, D. L. Clemens, M. A. Horwitz, Glutamine synthetase of Mycobacterium tuberculosis: extracellular release and characterization of its enzymatic activity. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 9342–9346 (1994).

161. T. C. Chou, P. Talalay, Generalized equations for the analysis of inhibitions of Michaelis-Menten and higher-order kinetic systems with two or more mutually exclusive and nonexclusive inhibitors. *Eur. J. Biochem.* **115**, 207–216 (1981).

162. S. L. Streicher, B. Tyler, Regulation of glutamine synthetase activity by adenylylation in the Gram-positive bacterium Streptomyces cattleya. *Proc. Natl. Acad. Sci. U.S.A.* **78**, 229–233 (1981).

163. A. Alissandratos, H. K. Kim, H. Matthews, Clostridium carboxidivorans Strain P7T Recombinant Formate Dehydrogenase Catalyzes Reduction of CO2 to Formate. *Appl. Environ. Microbiol.* **79**, 741–744 (2013).

164. J. Diruggiero, F. T. Robb, Expression and in vitro assembly of recombinant glutamate dehydrogenase from the hyperthermophilic archaeon Pyrococcus furiosus. *Appl. Environ. Microbiol.* **61**, 159–164 (1995).

165. T. Kigawa, Y. Muto, S. Yokoyama, Cell-free synthesis and amino acid-selective stable isotope labeling of proteins for NMR analysis. *J. Biomol. NMR.* **6**, 129–134 (1995).

166. C. Klammt, F. Löhr, B. Schäfer, W. Haase, V. Dötsch, H. Rüterjans, C. Glaubitz, F. Bernhard, High level cell-free expression and specific labeling of integral membrane proteins. *Eur. J. Biochem.* **271**, 568–580 (2004).

167. K. Ozawa, N. E. Dixon, G. Otting, Cell-free synthesis of 15N-labeled proteins for NMR studies. *IUBMB Life*. **57**, 615–622 (2005).

168. S. Reckel, S. Sobhanifar, B. Schneider, F. Junge, D. Schwarz, F. Durst, F. Löhr, P. Güntert, F. Bernhard, V. Dötsch, Transmembrane segment enhanced labeling as a tool for the backbone assignment of alpha-helical membrane proteins. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 8262–8267 (2008).

169. S. Sobhanifar, S. Reckel, F. Junge, D. Schwarz, L. Kai, M. Karbyshev, F. Löhr, F. Bernhard, V. Dötsch, Cell-free expression and stable isotope labelling strategies for membrane proteins. *J. Biomol. NMR.* **46**, 33–43 (2010).

170. N. M. Kredich, G. M. Tomkins, The enzymic synthesis of L-cysteine in Escherichia coli and Salmonella typhimurium. *J. Biol. Chem.* **241**, 4955–4965 (1966).

171. N. Michel-Reydellet, K. Calhoun, J. Swartz, Amino acid stabilization for cell-free protein synthesis by modification of the Escherichia coli genome. *Metabolic Engineering*. **6**, 197–203.

172. D. Schwarz, C. Klammt, A. Koglin, F. Löhr, B. Schneider, V. Dötsch, F. Bernhard, Preparative scale cell-free expression systems: new tools for the large scale preparation of integral membrane proteins for functional and structural studies. *Methods*. **41**, 355–369 (2007).

173. D. Schwarz, F. Junge, F. Durst, N. Frölich, B. Schneider, S. Reckel, S. Sobhanifar, V. Dötsch, F. Bernhard, Preparative scale expression of membrane proteins in Escherichia coli-based continuous exchange cell-free systems. *Nat Protoc*. **2**, 2945–2957 (2007).

174. O. Paliy, D. Bloor, D. Brockwell, P. Gilbert, J. Barber, Improved methods of cultivation and production of deuteriated proteins from E. coli strains grown on fully deuteriated minimal medium. *J. Appl. Microbiol*. **94**, 580–586 (2003).

175. M. Hochuli, T. Szyperski, K. Wüthrich, Deuterium isotope effects on the central carbon metabolism of Escherichia coli cells grown on a D2O-containing minimal medium. *J. Biomol. NMR*. **17**, 33–42 (2000).

176. W. K. Saur, D. T. Peterson, E. A. Halevi, H. L. Crespi, J. J. Katz, Deuterium isotope effects in the fermentation of hexoses to ethanol by Saccharomyces cerevisiae. II. A steady-state kinetic analysis of the isotopic composition of the methyl group of ethanol in an isotopic mirror fermentation experiment. *Biochemistry*. **7**, 3537–3546 (1968).

177. D. B. Northrop, Steady-state analysis of kinetic isotope effects in enzymic reactions. *Biochemistry*. **14**, 2644–2651 (1975).

178. A. Hattori, H. L. Crespi, J. J. Katz, Effect of side-chain deuteration on protein stability. *Biochemistry*. **4**, 1213–1225 (1965).

179. P. D. Sniegowski, P. J. Gerrish, R. E. Lenski, Evolution of high mutation rates in experimental populations of E. coli. *Nature*. **387**, 703–705 (1997).

180. P. Du, W. A. Kibbe, S. M. Lin, Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching. *Bioinformatics*. **22**, 2059–2065 (2006).

181. J. Slavič, I. Simonovski, M. Boltežar, Damping identification using a continuous wavelet transform: application to real data. *J Sound Vibration*. **262**, 291–307 (2003).

182. G. K. Smyth, Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol*. **3**, Article3 (2004).

183. Y. Benjamini, Y. Hochberg, Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Series B Stat Methodol*. **57**, 289–300 (1995).

184. O. Paliy, T. S. Gunasekera, Growth of E. coli BL21 in minimal media with different gluconeogenic carbon sources and salt contents. *Appl Microbiol Biotechnol*. **73**, 1169–1172 (2007).

185. Da Wei Huang, B. T. Sherman, R. A. Lempicki, Systematic and integrative analysis

of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. **4**, 44–57 (2008).

186. R. E. Bishop, B. K. Leskiw, R. S. Hodges, C. M. Kay, J. H. Weiner, The entericidin locus of Escherichia coli and its implications for programmed bacterial cell death. *J. Mol. Biol.* **280**, 583–596 (1998).

187. A. Conter, C. Menchon, C. Gutierrez, Role of DNA supercoiling and rpoS sigma factor in the osmotic and growth phase-dependent induction of the gene osmE of Escherichia coli K12. *J. Mol. Biol.* **273**, 75–83 (1997).

188. S. C. Brooks, Osmotic effects of deuterium oxide (heavy water) on living cells. *Science*. **86**, 497–498 (1937).

189. S. Scheiner, M. Čuma, Relative Stability of Hydrogen and Deuterium Bonds. *J. Am. Chem. Soc.* **118**, 1511–1521 (1996).

190. G. C. Kresheck, H. Schneider, H. A. Scheraga, The effect of D_2O on the thermal stability of proteins. Thermodynamic parameters for the transfer of model compounds from H_2O to D_2O. *J Phys Chem*. **69**, 3132–3144 (1965).

191. C. M. Armstrong, The Na/K pump, Cl ion, and osmotic stabilization of cells. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 6257–6262 (2003).

192. J. M. Santos, P. Freire, M. Vicente, C. M. Arraiano, The stationary-phase morphogene bolA from Escherichia coli is induced by stress during early stages of growth. *Mol Microbiol*. **32**, 789–798 (1999).

193. S. B. Farr, T. Kogoma, Oxidative stress responses in Escherichia coli and Salmonella typhimurium. *Microbiol. Rev.* **55**, 561–585 (1991).

194. Y. Zhang, M. Xiao, T. Horiyama, Y. Zhang, X. Li, K. Nishino, A. Yan, The multidrug efflux pump MdtEF protects against nitrosative damage during the anaerobic respiration in Escherichia coli. *J. Biol. Chem.* **286**, 26576–26584 (2011).

195. L. Nachin, U. Nannmark, T. Nyström, Differential roles of the universal stress proteins of Escherichia coli in oxidative stress resistance, adhesion, and motility. *J. Bacteriol.* **187**, 6265–6272 (2005).

196. Y. M. Efimova, S. Haemers, B. Wierczinski, W. Norde, A. A. van Well, Stability of globular proteins in H2O and D2O. *Biopolymers*. **85**, 264–273 (2007).

197. S. E. Chuang, F. R. Blattner, Characterization of twenty-six new heat shock genes of Escherichia coli. *J. Bacteriol.* **175**, 5242–5252 (1993).

198. M. Kitagawa, M. Miyakawa, Y. Matsumura, T. Tsuchido, Escherichia coli small heat shock proteins, IbpA and IbpB, protect enzymes from inactivation by heat and oxidants. *Eur. J. Biochem.* **269**, 2907–2917 (2002).

199. A. L. Fink, Chaperone-mediated protein folding. *Physiol. Rev.* **79**, 425–449 (1999).

200. K. Unno, T. Kishido, M. Morioka, S. Okada, N. Oku, Increased expression of Hsp70

for resistance to deuterium oxide in a yeast mutant cell line. *Biol. Pharm. Bull.* **26**, 799–802 (2003).

201. K. Unno, N. Hagima, T. Kishido, S. Okada, N. Oku, Deuterium-resistant algal cell line for D labeling of heterotrophs expresses enhanced level of Hsp60 in D2O medium. *Appl. Environ. Microbiol.* **71**, 2256–2259 (2005).

202. A. Zapun, D. Missiakas, S. Raina, T. E. Creighton, Structural and Functional Characterization of DsbC, a Protein Involved in Disulfide Bond Formation in Escherichia coli. *Biochemistry*. **34**, 5075–5089 (1995).

203. V. Bailly, W. G. Verly, The multiple activities of Escherichia coli endonuclease IV and the extreme lability of 5'-terminal base-free deoxyribose 5-phosphates. *Biochem. J.* **259**, 761–768 (1989).

204. J. J. Truglio, D. L. Croteau, B. Van Houten, C. Kisker, Prokaryotic nucleotide excision repair: the UvrABC system. *Chem. Rev.* **106**, 233–252 (2006).

205. P. C. Hanawalt, G. Spivak, Transcription-coupled DNA repair: two decades of progress and surprises. *Nat Rev Mol Cell Biol* (2008).

206. M. Freundlich, N. Ramani, E. Mathew, A. Sirko, P. Tsui, The role of integration host factor in gene expression in Escherichia coli. *Mol Microbiol.* **6**, 2557–2563 (1992).

207. K. K. Swinger, P. A. Rice, IHF and HU: flexible architects of bent DNA. *Curr Opin Struct Biol.* **14**, 28–35 (2004).

208. G. M. Dhavan, D. M. Crothers, M. R. Chance, M. Brenowitz, Concerted binding and bending of DNA by Eschericia coli integration host factor1. *J. Mol. Biol.* **315**, 1027–1037 (2002).

209. C. F. Higgins, C. J. Dorman, D. A. Stirling, L. Waddell, I. R. Booth, G. May, E. Bremer, A physiological role for DNA supercoiling in the osmotic regulation of gene expression in S. typhimurium and E. coli. *Cell*. **52**, 569–584 (1988).

210. A. Zhang, V. Derbyshire, J. L. Salvo, M. Belfort, Escherichia coli protein StpA stimulates self-splicing by promoting RNA assembly in vitro. *RNA*. **1**, 783–793 (1995).

211. W. Jiang, Y. Hou, M. Inouye, CspA, the major cold-shock protein of Escherichia coli, is an RNA chaperone. *J. Biol. Chem.* **272**, 196–202 (1997).

212. J. L. Yates, A. E. Arfsten, M. Nomura, In vitro expression of Escherichia coli ribosomal protein genes: autogenous inhibition of translation. *Proc. Natl. Acad. Sci. U.S.A.* **77**, 1837–1841 (1980).

213. C. K. Tang, D. E. Draper, Evidence for allosteric coupling between the ribosome and repressor binding sites of a translationally regulated mRNA. *Biochemistry* (1990).

214. D. E. Agafonov, A. S. Spirin, The ribosome-associated inhibitor A reduces translation errors. *Biochem. Biophys. Res. Commun.* **320**, 354–358 (2004).

215. J. Ling, D. Söll, Severe oxidative stress induces protein mistranslation through impairment of an aminoacyl-tRNA synthetase editing site. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 4028–4033 (2010).

216. M. J. Lukey, A. Parkin, M. M. Roessler, B. J. Murphy, J. Harmer, T. Palmer, F. Sargent, F. A. Armstrong, How Escherichia coli Is Equipped to Oxidize Hydrogen under Different Redox Conditions. *J. Biol. Chem.* **285**, 3928–3938 (2010).

217. P. M. Vignais, B. Billoud, Occurrence, classification, and biological function of hydrogenases: an overview. *Chem. Rev.* **107**, 4206–4272 (2007).

218. K. Trchounian, A. Trchounian, Escherichia coli multiple [Ni–Fe]-hydrogenases are sensitive to osmotic stress during glycerol fermentation but at different pHs. *FEBS Lett.* **587**, 3562–3566 (2013).

219. Y. P. Chao, R. Patnaik, W. D. Roof, R. F. Young, J. C. Liao, Control of gluconeogenic growth by pps and pck in Escherichia coli. *J. Bacteriol.* **175**, 6939–6944 (1993).

220. M. Emmerling, M. Dauner, A. Ponti, J. Fiaux, M. Hochuli, T. Szyperski, K. Wüthrich, J. E. Bailey, U. Sauer, Metabolic flux responses to pyruvate kinase knockout in Escherichia coli. *J. Bacteriol.* **184**, 152–164 (2002).

221. H. L. Kornberg, The role and control of the glyoxylate cycle in Escherichia coli. *Biochem. J.* **99**, 1–11 (1966).

222. G. L. Lorca, A. Ezersky, V. V. Lunin, J. R. Walker, S. Altamentova, E. Evdokimova, M. Vedadi, A. Bochkarev, A. Savchenko, Glyoxylate and pyruvate are antagonistic effectors of the Escherichia coli IclR transcriptional regulator. *J. Biol. Chem.* **282**, 16476–16491 (2007).

223. C. Yanofsky, V. Horn, Role of regulatory features of the trp operon of Escherichia coli in mediating a response to a nutritional shift. *J. Bacteriol.* **176**, 6245–6254 (1994).

224. K. B. Storey, *Functional Metabolism: Regulation and Adaptation* (Wiley, 2004).

225. F. Ozsolak, P. M. Milos, RNA sequencing: advances, challenges and opportunities. *Nat. Rev. Genet.* **12**, 87–98 (2011).

226. T. J. Griffin, S. P. Gygi, T. Ideker, B. Rist, J. Eng, L. Hood, R. Aebersold, Complementary profiling of gene expression at the transcriptome and proteome levels in Saccharomyces cerevisiae. *Mol. Cell Proteomics.* **1**, 323–333 (2002).

227. C. Murie, O. Woody, A. Y. Lee, R. Nadon, Comparison of small n statistical tests of differential expression applied to microarrays. *BMC Bioinformatics.* **10**, 45 (2009).

228. A. R. Aricescu, W. Lu, E. Y. Jones, A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 1243–1250 (2006).

229. S. Chaudhary, J. E. Pak, F. Gruswitz, V. Sharma, R. M. Stroud, Overexpressing

human membrane proteins in stably transfected and clonal human embryonic kidney 293S cells. *Nat Protoc*. **7**, 453–466 (2012).

230. A. Dutta, K. Saxena, H. Schwalbe, J. Klein-Seetharaman, Isotope labeling in mammalian cells. *Methods Mol. Biol.* **831**, 55–69 (2012).

231. J. Klein-Seetharaman, P. J. Reeves, M. C. Loewen, E. V. Getmanova, J. Chung, H. Schwalbe, P. E. Wright, H. G. Khorana, Solution NMR spectroscopy of [alpha - 15N]lysine-labeled rhodopsin: The single peak observed in both conventional and TROSY-type HSQC spectra is ascribed to Lys-339 in the carboxyl-terminal peptide sequence. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 3452–3457 (2002).

232. A. B. Patel, E. Crocker, P. J. Reeves, E. V. Getmanova, M. Eilers, H. G. Khorana, S. O. Smith, Changes in interhelical hydrogen bonding upon rhodopsin activation. *J. Mol. Biol.* **347**, 803–812 (2005).

233. J. Klein-Seetharaman, N. V. K. Yanamala, F. Javeed, P. J. Reeves, E. V. Getmanova, M. C. Loewen, H. Schwalbe, H. G. Khorana, Differential dynamics in the G protein-coupled receptor rhodopsin revealed by solution NMR. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 3409–3413 (2004).

234. K. Werner, I. Lehner, H. K. Dhiman, C. Richter, C. Glaubitz, H. Schwalbe, J. Klein-Seetharaman, H. G. Khorana, Combined solid state and solution NMR studies of alpha,epsilon-15N labeled bovine rhodopsin. *J. Biomol. NMR*. **37**, 303–312 (2007).

235. K. Werner, C. Richter, J. Klein-Seetharaman, H. Schwalbe, Isotope labeling of mammalian GPCRs in HEK293 cells and characterization of the C-terminus of bovine rhodopsin by high resolution liquid NMR spectroscopy. *J. Biomol. NMR*. **40**, 49–53 (2008).

236. J. W. Lustbader, S. Birken, S. Pollak, A. Pound, B. T. Chait, U. A. Mirza, S. Ramnarain, R. E. Canfield, J. M. Brown, Expression of human chorionic gonadotropin uniformly labeled with NMR isotpes in Chinese hamster ovary cells: An advance toward rapid determination of glycoprotein structures. *J. Biomol. NMR*. **7**, 295–304 (1996).

237. K. Shindo, K. Masuda, H. Takahashi, Y. Arata, I. Shimada, Letter to the Editor: Backbone 1H, 13C, and 15N resonance assignments of the anti-dansyl antibody Fv fragment. *J. Biomol. NMR*. **17**, 357–358 (2000).

238. T. A. Egorova-Zachernyuk, G. J. C. G. M. Bosman, W. J. DeGrip, Uniform stable-isotope labeling in mammalian cells: formulation of a cost-effective culture medium. *Appl Microbiol Biotechnol*. **89**, 397–406 (2010).

239. J. Goncalves, M. Eilers, K. South, C. A. Opefi, P. Laissue, P. J. Reeves, S. O. Smith, Magic angle spinning nuclear magnetic resonance spectroscopy of G protein-coupled receptors. *Meth. Enzymol.* **522**, 365–389 (2012).

240. M. Schneider, I. W. Marison, U. von Stockar, The importance of ammonia in mammalian cell culture. *J. Biotechnol.* **46**, 161–185 (1996).

241. G. Zhang, V. Gurtu, S. R. Kain, An enhanced green fluorescent protein allows sensitive detection of gene transfer in mammalian cells. *Biochem. Biophys. Res. Commun.* **227**, 707–711 (1996).

242. D. D. Oprian, R. S. Molday, R. J. Kaufman, H. G. Khorana, Expression of a synthetic bovine rhodopsin gene in monkey kidney cells. *Proc. Natl. Acad. Sci. U.S.A.* **84**, 8874–8878 (1987).

243. R. S. Molday, D. MacKenzie, Monoclonal antibodies to rhodopsin: characterization, cross-reactivity, and application as structural probes. *Biochemistry*. **22**, 653–660 (1983).

244. P. J. Reeves, J.-M. Kim, H. G. Khorana, Structure and function in rhodopsin: a tetracycline-inducible system in stable mammalian cell lines for high-level expression of opsin mutants. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 13413–13418 (2002).

245. P. J. Reeves, N. Callewaert, R. Contreras, H. G. Khorana, Structure and function in rhodopsin: High-level expression of rhodopsin with restricted and homogeneous N-glycosylation by a tetracycline-inducible N-acetylglucosaminyltransferase I-negative HEK293S stable mammalian cell line. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 13419–13424 (2002).

Curriculum vitae

Christian Opitz, PhD
Dorfstr. 17 | 79595 Rümmingen (Germany) | Phone: +49 7621 1628957
christian.opitz@unibas.ch | christian@opitz.science

Date and place of birth:	March, 7th, 1984, Quedlinburg, Germany
Nationality:	German
Marital status:	married, one child

Education and qualifications

05/2016–present	Postdoctoral researcher, group of Stephan Grzesiek Biozentrum, University of Basel (Switzerland)
01/2012–04/2016	PhD in biophysics, Prof. Stephan Grzesiek Biozentrum, University of Basel (Switzerland)
09/2010–12/2011	Research work in the lab of Prof. Sebastian Hiller Biozentrum, University of Basel (Switzerland)
2009–2010	Diploma thesis *Studies on the dynamics and expression of the Rubella virus capsid protein in living cells* (Prof. Uwe Gerd Liebert) University of Leipzig (Germany)
2004–2009	Studies in biology Specialization in molecular virology, biophysics, biochemistry, plant and behavioral physiology University of Leipzig (Germany)

2003–2004	Dual studies in medical information management German Cancer Research Center Heidelberg and Baden-Wuerttemberg Cooperative State University (Heidelberg, Germany)

Course work

11/2015	Patenting and Spin-off workshop
08/2012	EMBO practical course on *Multidimensional NMR in Structural Biology*
01/2011	NCCR course on *Reaction Kinetics of Biological Macromolecules*

Awards

2017	Best poster award, "Advanced Isotopic Labeling Methods for Integrated Structural Biology" (AILM)

Languages

German (mother tongue), English (highly proficient),

Basic knowledge of Swedish and Russian

Publications

T. Raschle, P. Rios Flores, **C. Opitz**, D. J. Müller, S. Hiller, Monitoring Backbone Hydrogen-Bond Formation in β-Barrel Membrane Protein Folding. *Angew Chem Int Ed Engl.* **55**, 5952–5955 (2016).

S. Isogai, X. Deupi, **C. Opitz**, F. M. Heydenreich, C.-J. Tsai, F. Brueckner, G. F. X. Schertler, D. B. Veprintsev, S. Grzesiek, Backbone NMR reveals allosteric signal transduction networks in the β1-adrenergic receptor. *Nature*. **530**, 237–241 (2016).

C. Opitz, S. Isogai, S. Grzesiek, An economic approach to efficient isotope labeling in insect cells using homemade 15N-, 13C- and 2H-labeled yeast extracts. *J. Biomol. NMR*. **62**, 373–385 (2015).

R. Goss, **C. Opitz**, B. Lepetit, C. Wilhelm, The synthesis of NPQ-effective zeaxanthin depends on the presence of a transmembrane proton gradient and a slightly basic stromal side of the thylakoid membrane. *Planta*. **228**, 999–1009 (2008).

References

Professor Stephan Grzesiek

Stephan.grzesiek@unibas.ch | Biozentrum, University of Basel | Klingelbergstrasse 50/70, CH - 4056 Basel

Professor Timm Maier (Co-referee PhD thesis)

Timm.maier@unibas.ch | Biozentrum, University of Basel | Klingelbergstrasse 50/70, CH - 4056 Basel