Development of a Microfluidics-Based Screening Assay for the High-Throughput Directed Evolution of Artificial Metalloenzymes

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Jaicy Vallapurackal

Basel, 2023

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von

Erstbetreuer: Prof. Dr. Thomas R. Ward

Zweitbetreuer: Prof. Dr. Florian P. Seebeck

Externer Experte: Prof. Dr. Gerard Roelfes

Basel, den 22.06.2021

Prof. Dr. Marcel Mayor

Dekan

“Any sufficiently advanced technology is indistinguishable from magic.”
— *Arthur C. Clarke (1917-2008)*

To my parents
പപ്പാക്കും മമ്മിക്കും

ACKNOWLEDGEMENTS

First and foremost, I would like to thank *PROF. DR. THOMAS R. WARD* for accepting me as a PhD student in his group. I had the opportunity to work on challenging and fascinating projects, through which I was able to learn a lot and outgrow myself as a scientist. I will always value the opportunities that you made possible, and your guidance and mentoring, that helped me grow, both professionally as well as personally.

Moreover, I would sincerely like to thank *PROF. FLORIAN P. SEEBECK* for agreeing to co-referee this dissertation and *PROF. GERARD ROELFES* for accepting to be the external expert.

My special thanks goes to *ARIANE STUCKI*, with whom I have been collaborating over the past 3 years. Only rarely one is lucky enough to find a collaboration partner, with whom you can work so well. I am extremely grateful for both, our scientific collaboration and our friendship! Your ever so patient nature (apart from lunchtime hangryness) and support have helped me to get through the worst and best moments of my PhD studies, and for that I am indebted to you. I cannot imagine how I would have done it without you.

Furthermore, I had the pleasure to be involved in numerous collaborations and would like to express my gratitude towards the groups of *PROF. PETRA S. DITTRICH, DR. MARKUS JESCHEK, PROF. SVEN PANKE, PROF. CORNELIA PALIVAN* and *PROF. WOLFGANG MEIER*, and *PROF. ANDREW S. BOROVIK*, as well as the PhD and PostDoc collaborators *ARIANE STUCKI, TOBIAS VORNHOLT, DR. PHILIPP ROTTMANN, STEFANO DI LEONE*, and *DR. LISA OLSHANSKY*.

My huge thanks goes to the whole *DrEAM* team, including the past and present members, who helped me to even further broaden my knowledge. I would especially like to mention *RYAN*, who had this innate curiosity that he would share with everyone. From picking up one or the other thing about chemistry and biology, to making the perfect sourdough starter, it has been a big joy to be your labmate.

ALI, thank you very much for your support especially over the past year! I could always drop by your lab with my scientific problems, which you were able to help me with creative solutions, due to your widely facetted knowledge. *ALINA* and *SNIZHANA*, thank you so much for your friendship and the occasional coffee breaks and walks in the Lange Erle –those breaks were worth gold, especially in these unprecedented times! My thanks also go out to Daniel and Steven, my former students, for their contributions towards the project, and *ALAIN* for the collaboration and all the valuable discussions. *JULIANE*, you have taught me so many techniques in the biolab, starting from

growing cultures to Western blotting. Your guidance and patience have led me through my numerous endeavours in biology. And of course, I will never forget the fun we had in Paris! *NICO* and *ROBIN*, heroes of late evenings and weekends, thanks for the occasional emergency assistance! *JO* and *ISABEL*, you have been fantastic neighbors in the open office. I am so grateful for our *"Tratschecke"* and discussions on science and life in general. At this point I would also like to express my gratitude towards the older generation of the Ward-group, who I looked up to ever since I was a Master student. Firstly, thanks *MARTINA* for your mentorship (still!). And thanks *RAPHAEL*, *HENDRIK*, *SASCHA*, *MICHELA*, *VINCENT*, *CHRISTIAN*, *TILLMANN* for being great role models and for all the fun we had, be it in or outside of the lab. Thanks *JONAS*, *FABIAN*, *FADRI*, *VALERIO*, *YASU*, *JINGMING*, *JOHANNES*, *ISABEL*, *JOAN*, *HOLLY*, *YOANN*, *DONGPING*, *IORI*, *RYO*, *CORENTIN*, *ALAIN*, *ALINA*, *SNIZHANA*, and *VALERIE*, and everyone I might have forgotten for being part for the group. As one of you once told me: "I'm sorry, that this part of the journey is so stressful, but I'm more than glad that I'm not alone and can count on you!"

Research would not be possible without a great environment. That was provided by administrative staff, the Werkstatt Team, the single cell facility at the DBSSE, the FACS Core facility at the Biozentrum (special shoutout to *STELLA* and *JANINE* who patiently measured our manifold DE samples), the NMR support and IT support. A big thanks especially to *ISA*, *ESTHI* and *CLAUDIA*, for all your help with the administrative side of the studies and the kind words! For the past three years, I have also been actively involved in scientific outreach. Here, I would like to thank the NCCR-MSE, especially *DR. RALF STUTZKI*, for the memorable events, and the opportunity to discuss various other aspects of science. Especially filming the TechThursday videos was a beautiful way to relax as well as engage with the community and learn about new topics.

I am incredibly grateful to my friends inside and outside of the Department who I can always count on. I would like to thank especially *BASILIUS* and *SIMON* for their great support with python scripts and data analysis at any time of the day. I would also like to name the BaseCamp Team, the Pint of Science organizers across Switzerland and all the PCC members for your friendship and fun times. Moreover, I would like to thank *ALEX*, *DOMINIQUE*, *ANDREA* and *ISABEL*, for your friendship. Particularly the countless strolls along the Rhine during the pandemic were a great relief during the stressful times, your friendship is worth far more than words could contain!

I would not be here without the support of my family. I would like to thank my siblings, *JAEY* and *JAIMY*, for always having my back and enduring countless presentations in advance to the real one and their honest feedback on anything I show them. Last but not least, my greatest appreciation goes to my parents, *SHAJI* and *GRACY*, who always have been my rock in turbulent waters and created opportunities, where there were none in the beginning. I love you!

PREFACE

The present PhD thesis summarizes the scientific work conducted in the research group of Prof. Dr. Thomas R. Ward during the years 2016–2021.

Research in the Ward group is focused on the development and optimization of artificial metalloenzymes with non-natural activities. These hybrid catalysts, resulting from an incorporation of a metal–containing cofactor within a protein or DNA scaffold, and can be optimized by either chemical or genetic means.

The main part of this thesis deals with the genetic optimization of such systems and the development of higher throughput screening assays to facilitate the process. First attempts dealt with the development of a selection-based assay relying on the Carroll rearrangement (Chapter 2.6). Following, more high-throughput assays such as screening of cells relying on a fluorescent reporter protein (Chapter 3) or the screening of activity by an agar plate screening assay were pursued (Chapter 4.2).

The main part of the thesis focuses on the method development of an ultrahigh-throughput screening platform for the *in vivo* directed evolution of artificial metalloenzymes using droplet microfluidics. The combination of ArMs and droplet microfluidics, can be a powerful tool for propelling directed evolution-based research forward. Systematic and high-throughput screening of ArMs *in vivo* using double emulsions could allow the screening of a much bigger sequence space, which is, to date, challenging. Identifying cooperative effects to improve catalysis or even remodelling whole enzymes to achieve new-to-nature reactivities are only two potential examples. Reactions based on ArMs could ultimately provide aqueous, environmentally friendly reaction pathways for industrial applications. Additionally, such big data sets could also be used as an input for machine learning applications, to further study active site plasticity, reaction pathways, or even protein-folding mechanisms.

The developed method was then applied to libraries of different types and sizes, and recent findings of these screenings are highlighted in the fourth chapter.

During the time in the research group of Prof. Dr. Ward, a deeper knowledge in molecular biology, especially library design, high-throughput screening using different approaches, microfluidic method development and fluorescence activated cell sorting (FACS), and the use of different sequencing techniques was garnered.

LIST OF PUBLICATIONS

List of publications that were published or are to be submitted:

- Vallapurackal, J.; Stucki, A.; Liang, A. D.; Dittrich, P. S.; Ward, T. R. *"Ultrahigh-Throughput Screening of an Artificial Metalloenzyme using an Ultrahigh-Throughput Double Emulsion Screening Assay"*
  Angew. Chem. Int. Ed. ***2022****, 61, e2022073.*

- Schwizer F.; Baiyoumy A.; Vallapurackal J.; Schwizer F.; Heinisch T.; Kardashliev T.; Held, M.; Panke S.; Ward T. R. *"Directed Evolution of a Surface-Displayed Artificial Allylic Deallylase Relying on a GFP Reporter Protein."*
  ACS Catal. **2021**, *11* (48), 1070–510712.

- Leone, S. D.; Vallapurackal J.; Avsar, S. Y.; Kyropolou, M.; Ward, T. R.; Palivan, C. G.; Meier, W. *"Expanding the Potential of Solvent-Assisted Method to Create Bio-Interfaces From Amphiphilic Block Copolymers."*
  Biomacromolecules, **2021**, *22* (7), 3005–3016.

- Vallapurackal J.; *"Towards the Directed Evolution of Artificial Metalloenzymes."*
  Chim. Int. J. Chem. **2021**, *75* (4), 257–260.

- Stucki, A.; Vallapurackal, J.; Ward, T. R.; Dittrich, P. S. "*Droplet Microfluidics and Directed Evolution of Enzymes: An Intertwined Journey."*
  Angew. Chemie Int. Ed. **2021**, *60* (46), 24368–24387.

- Heinisch, T.; Schwizer, F.; Garabedian, B.; Csibra, E.; Jeschek, M.; Vallapurackal, J.; Pinheiro, V. B.; Marlière, P.; Panke, S.; Ward, T. R. *"E. Coli Surface Display of Streptavidin for Directed Evolution of an Allylic Deallylase.*"
  Chem. Sci. **2018**, *9* (24), 5383–5388.

- Olshansky, L.; Huerta-Lavorie, R.; Nguyen, A. I.; Vallapurackal, J.; Furst, A.; Tilley, T. D.; Borovik, A. S. "*Artificial Metalloproteins Containing Co_4O_4 Cubane Active Sites."*
  J. Am. Chem. Soc. **2018**, *140* (8), 2739–2742.

CONTRIBUTIONS

Following people contributed with their work to this thesis:

MSc. Ariane Stucki (Dittrich group, ETH Zürich):
Ariane Stucki designed, fabricated and operated the microfluidic device used for encapsulation of the *E. coli* in Chapter 4.3.2. She also performed most of the flow cytometry measurements of double emulsions and was involved in the data analysis of the sequencing results presented in Chapter 4.5.

MSc. Tobias Vornholt (Jeschek group, ETH Zürich)
Tobias Vornholt provided the 400-variants library and the pSC101 plasmid bearing the mNectarine and GFP genes used in Chapter 4.5.

Dr. Alexandria Deliz Liang (Ward group, University of Basel)
Alexandira Deliz Liang helped in the design of the combinatorial variant library purchased from Twist Biosciences mentioned in Chapter 4.5.

Dr. Markus Jeschek (ETH Zürich)
Markus Jeschek guided the Illumina sequencing approach and reviewed the design of the corresponding primers mentioned in Chapter 4.5.

Dr. Fabian Schwizer (Ward group, University of Basel)
Fabian Schwizer cloned the Sav library used in Chapter 3.3 and performed the *in vivo* screening described in Chapter 3.3.

MSc. Alain Baiyoumy (Ward group, University of Basel)
Alain Baiyoumy performed the *in vivo* and *in vitro* experiments described in Chapter 3.3.

Dr. Tsvetan Kardashliev (Panke group, ETH Zürich):
Tsvetan Kardashliev designed and cloned the DmpR/GFP reporter system in Chapter 3.3.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

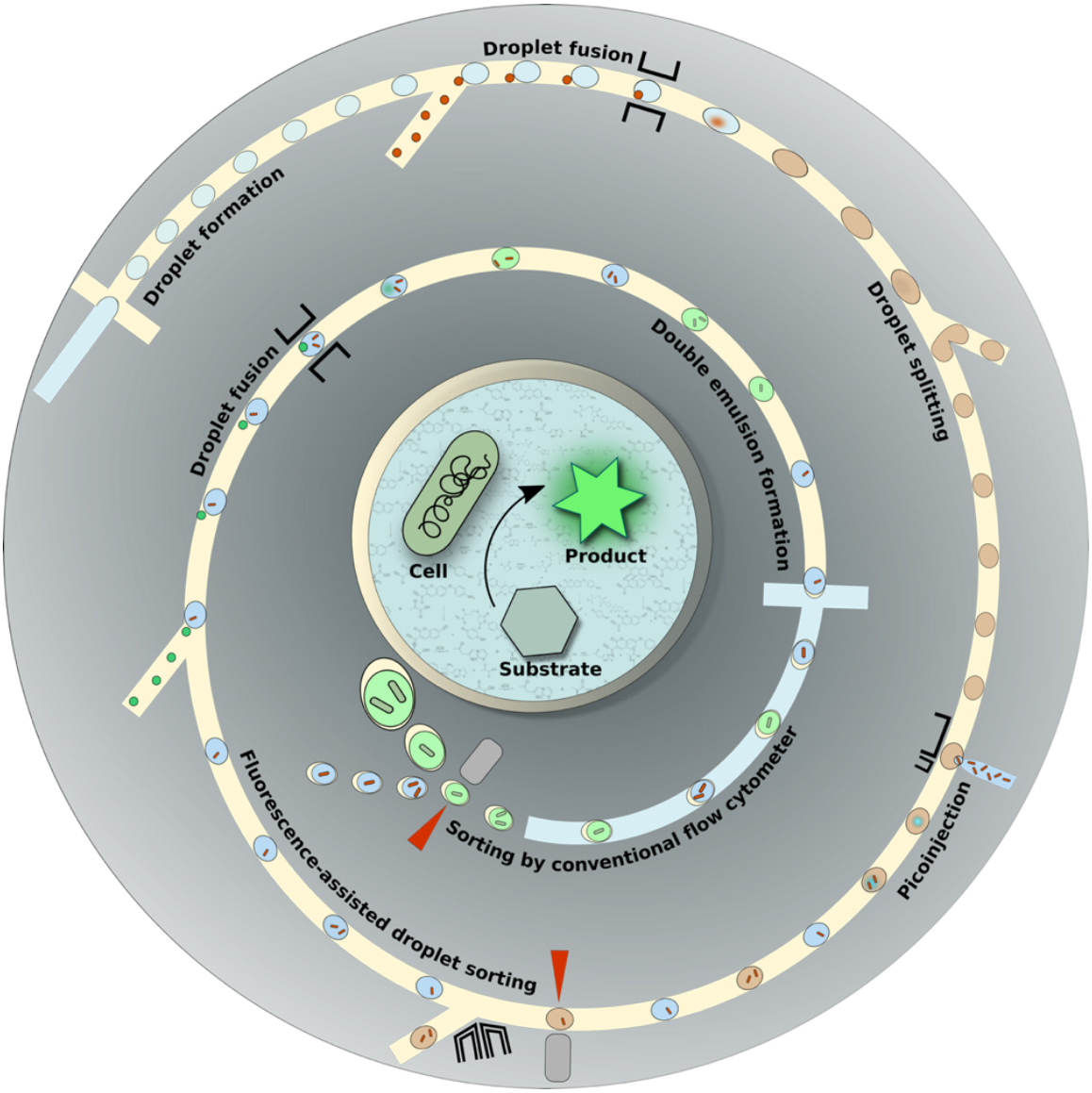
δ	Chemical shift
AADS	Absorbance-activated droplet sorting
ACN	Acetonitrile
ADAse	Artificial Deallylase
AFEST	*Archeoglobus fulgidus* thermophilic esterase
AHEG	anthracentylmethyl hexaethylene glycol
αHL	α-hemolysin
ArMs	Artificial metalloenzymes
ATA	*Arthrobacter sp.* transaminase
B4F	Biotin-4-fluorescein
BS^3	Bis(sulfosuccinimidyl) suberate
BSA	Bovine serum albumin
cfu	Colony forming unit
CHAO	cyclohexylamine oxidase
conc.	Concentration
conv.	Conversion
CpI	Clostridium pasteurianum PpAAR pichia pastoris
CSR	Compartmentalized self-replication
DNA	Deoxyribonucleic acid
DCM	Dichloromethane
DE	Double emulsion
DEP	Dielectrophoresis
DHFR	Dihydrofolate reductase
DiBT-IMMS	desorption electrospray ionization coupled with ion mobility mass spectrometry imaging
DMDS	dual-channel microfluidic droplet screening system
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
dNDPs	Deoxyribose nucleoside diphosphates
dNTPs	Deoxyribose nucleoside triphosphates
DpnI	Diplococcus pneumoniae restriction enzyme
E.coli	Escherichia coli
ee	Enantiomeric excess
epPCR	Error prone polymerase chain reaction
ESI-MS	Electron-spray ionization mass
EtOAc	Ethylacetate
EtOH	Ethanol
FACS	Fluorescence-Activated Cell Sorting
FAC	Fluorescent correlation spectroscopy
FADS	Fluorescence activated droplet sorting
FBS	Free biotin binding sites
FESEM	Field Emission Scanning Electron Microscope
FI	Fluorescence intensity
FSC	Forward scatter
GC	Gas chromatography
GFP	Green fluorecent protein
Gox	Glucose oxidase
GSH	Glutathione
GSSG	Glutathione disulfide
h	Hours
hCA	Human carbonic anhydrase
HMF	5-hydroxymethylfurfural
HOV-I	Hoveyda-Grubbs Catalyst 2nd generation
HPLC	High performance lquid chromatography
HRP	Horseradish peroxidase

HsAf	Horse spleen apoferritin
HTS	High throughput screening
IA	Inner aqueous phase
ICP-MS	Inductively-coupled plasma mass spectroscopy
IDH	isocitrate dehydrogenase
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IRED	Imine reductase
ISS	International Sapce Agency
IVC	In vitro compartmentalization
ivTT	In vitro transcrition and translation
LB	Lysogeny broth / Luria-Bertani medium
LC	Liquid chromatography
LC-MS	Liquid chromatography coupled mass spectroscopy
Lpp	Lipoprotein
MADS	Mass activated drolet sorting
MALDI-MS	Matrix-assisted laser desorption/ionization mass spectroscopy
MAO	Monoamine oxidase
MD	Molecular dynamics
MeOH	Methanol
Metathesase	Artificial metathesase
min	Minutes
MOPS	3-(N-morpholino)propanesulfonic acid
MS	Mass spectroscopy
MTPs	Multititer plate
NAD+	Nicotinamide adenine dinucleotide
NADH	Nicotinamide adenine dinucleotide hydride (reduced form)
native PAGE	Native polyacrylamide gel electrophoresis
NDK	nucleoside diphosphate kinase
NGS	Next generation sequencing
NMM	N-methylmorpholine
NMR	Nuclear magnetic resonance
OA	Outer aqueous phase
OD	Optical density
OmpA	Outer membrane protein A
OP	Oil phase
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PCR	Polymerase chain reaction
Pd/C	Palladium-saturated charcoal
PDB	Protein Data Bank
PDMS	Polydimethylsiloxane
pH	Potential of hydrogen
PON1	Paraoxonase 1
ppm	Parts per million
prot-SNPs	Silica nanoparticles protected in an organosilane layer
PTE	phosphotriesterase
PVA	Polyvinyl alcohol
rac.	Racemic mixture
RNA	Ribonucleic acid
rpm	Rotations per minute
RT	Room temperature
s.d.	Standard deviation
Sav	Streptavidin
scdSav	single chain dimeric Sav
SDS	Sodium dodecylsulphate
SDS-PAGE	Sodium dodecylsulphate polyacrylamide gel electrophoresis
SE	Single emulsion
SSC	Side scatter
Sub.	Substrate
SUMO	Small ubiquitin-related modifier
TBST	Tris-buffered saline + 0.1% Tween 20
TEMED	Tetramethylethylenediamine

TFA	Trifluoroacetic acid
TFAA	Trifluoroacetic anhydride
TMS	Tetramethylsilane
TNA	a-L-threofuranosyl nucleic acid
TOF	Turnover frequency
TON	Turnover number
TOPO	Topoisomerase I
Tris	Tris(hydroxymethyl)aminomethane
UPC^2	Ultraperformance convergence chromatography
UPLC	Ultra-performance liquid chromatography
w/o	Water in oil
w/o/w	Water in oil in water
wt	Wild type
X-ray	X-ray crystallography
ZYP	ZYP-5052 medium for auto-induction
PFP-TFA	Pentafluoropheny trifluoroacetate

1

Droplet Microfluidics & Directed Evolution of Enzymes



The following section has been published in
Angew. Chem. Int. Ed. ***2021****, DOI.*
doi: 10.1002/anie.202016154

Droplet Microfluidics and Directed Evolution of Enzymes: an Intertwined Journey

Ariane Stucki[#[a,c]], Jaicy Vallapurackal[#[b,c]], Thomas R. Ward[*[b,c]], Petra S. Dittrich[*[a,c]]

[a] Department of Biosystems Science and Engineering, ETH Zurich, Mattenstrasse 26, CH-4058 Basel, Switzerland. E-mail: petra.dittrich@bsse.ethz.ch

[b] Department of Chemistry, University of Basel, Mattenstrasse 24a, CH-4058 Basel, Switzerland. E-mail: thomas.ward@unibas.ch

[c] National Competence Center in Research (NCCR), Molecular Systems Engineering, Basel, Switzerland

[#] These authors contributed equally.

KEYWORDS: directed evolution · microfluidics · enzymes · droplet microfluidics

1.1 ABSTRACT

Evolution is essential to the appearance of complexity and ultimately Life. It relies on the propagation of the properties, traits and characteristics that allow an organism to survive in a challenging environment. It is evolution that shaped our world over about four billion years by slow and iterative adaptation. While natural evolution based on selection is slow and gradual, directed evolution allows the fast and streamlined optimization of a phenotype under selective conditions. The potential of directed evolution for the discovery and optimization of enzymes is mostly limited by the throughput of the tools and methods available for screening. Over the past twenty years, versatile tools based on droplet microfluidics have been developed to address the need for higher throughput. In this review, we provide a chronological overview of the intertwined development of microfluidics droplet-based compartmentalization methods and *in vivo* directed evolution of enzymes.

1.2 INTRODUCTION

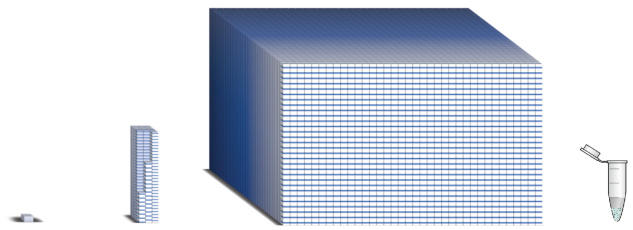
Enzymes, nature's privileged catalysts, were optimized for a specific biological purpose and evolved over thousands of generations by natural selection. Lowering reaction barriers to selectively enable and accelerate certain reactions is a key characteristic of enzymes. But as the enzymes' natural activities are often insufficient to meet the needs of mankind, artificial selection and screening have gained importance. Starting from the breeding of crops and domestication of animals for sustaining early populations, it matured to directed evolution in order to improve natural systems and introduce new-to-nature reactions for Life-Sciences and other applications.

The 2018 Nobel Prize in Chemistry was awarded for efforts in the development of directed evolution to Frances H. Arnold for the directed evolution of enzymes, and George P. Smith and Sir Gregory P. Winter for the phage display of peptides and antibodies. Directed evolution allows to alter and thus potentially improve biological activities by genetic means, an approach generally faster and with better control than natural selection. Accordingly, the study of these enzymes is of high importance to scientific advancement and holds great industrial potential as it allows the evolution of alternative or new reaction pathways in a streamlined fashion. This approach thus provides environmentally friendly pathways to valorize enzymes as an alternative to the more traditional chemistry toolbox.[1]

Applying directed evolution consists of three steps: *i)* to iteratively mutate (create genetic diversity), *ii)* screen (optimize for a desired property) and *iii)* choose (pick the best performing variant). If the protein of interest is well characterized, focused mutagenesis strategies can be implemented, followed by lower throughput screening.[2] In a pioneering study, Arnold and

coworkers highlighted the potential of directed evolution using subtilisin E. By screening about 4000 colonies, they evolved a variant capable of hydrolyzing a peptide substrate with 256-fold higher efficiency than wild-type in 60% dimethylformamide (DMF).[3] Since then, numerous *in vitro* and *in vivo* studies based on focused libraries in microtiter plates (MTPs) have been reported.[4] Selected examples include *i)* the directed evolution of sortase A to improve its robustness and activity by focused loop engineering and head-to-tail backbone cyclization,[5] *ii)* the directed evolution of enantiospecific enzymes,[6–8] *iii)* the directed evolution of P450 for various applications,[9–12] and, more recently, *iv)* the directed evolution of a *de novo* designed retro-aldolase,[13] of a metalloenzyme for enantiospecific ester-hydrolysis designed from short peptides,[14] and of a metalloenzyme for olefin metathesis using an expanded nitrobindin variant.[15] Lately, directed evolution finds also increased use in the biotechnological field: e.g. the process and enzyme engineering approach applied to galactose oxidase for the biocatalytic transformation of 5-hydroxymethylfurfural (HMF), a valuable building block in the synthesis of materials from renewable resources.[16] Apart from MTP-assays, another medium-throughput approach is the use of agar plate-based screening assays, which was illustrated with the directed evolution of transaminases as biocatalysts for chiral amine synthesis.[17] Enclosing the enzymatic reaction within cells or immobilizing fluorescent products on the cell surface is yet another strategy to increase the throughput and was applied to several systems, such as the evolution of a P450 monooxygenase.[18]

If the structure-activity relationships of the protein are poorly understood, more mutants may need to be screened to achieve a targeted phenotype. This is often achieved through a more thorough mutagenesis campaign of the protein and leads therefore to an exponential growth in the number of variants to be screened.[19]



randomized positions	1	2	4	4
screening effort	94	3066	3’141’251	3’141’251
96-well plates	1	32	32’722	50 µL of droplets
screening time	1 week	4 weeks	~79 years	1 week

Figure 1. Schematic representation of the screening effort per mutated position using an NNK library. An NNK library at one position has 32 possible codons encoding for the twenty amino acids. This corresponds to a screening effort of 94 colonies to achieve a theoretical library coverage of 95%. This effort increases exponentially if two or more positions are screened simultaneously. Screening four positions would require about 80 years, considering that eight 96-well plates are screened per week. In comparison, screening the same library in double emulsions using microfluidic tools would require about one week of work.[19]

Let us consider an example whereby four positions are simultaneously randomized. Using conventional screening assays based on MTPs may require over 80 years (roughly 20 PhDs!) of manual screening and appreciable amounts of material such as screening buffers (> 100 L) or costly catalyst solutions (>1 L).[20] In contrast, the same screening using double emulsions and Fluorescence-Activated Cell Sorting (FACS) could be performed in roughly a week by a single operator with substantially lower amounts of material (Figure 1).

Additionally, to enable large screening efforts, optical readouts such as color, fluorescence or luminescence are essential. In general, the industrially-relevant target products lack readily detectable phenotypes. In such cases, substrate analogues with a fluorescent, luminescent or colorimetric readout that correlates with the enzyme activity need to be implemented.

One of the main requirements in directed evolution is linking the activity of a target enzyme (i.e. the phenotype) to its genetic information (i.e. the genotype), which is essential for screening, selection and ultimately evolution campaigns.[21,22] To address this challenge, different strategies, such as compartmentalizing the enzymatic reaction within/on cells or immobilizing fluorescent products on the cell surface, have been explored and are described in detail in other reviews.[23,24] These strategies opened the way to high-throughput analysis methods such as FACS. FACS devices have gained increasing interest since their initial development and the first instrument commercialization in the late 1960s-1970s.[25,26] The development of microfluidic devices for fluorescence-based particle- or cell-sorting using negative dielectrophoresis (DEP) contributed to the early advancement of such technologies.[27] Other fluorescence-based methods such as fluorescent correlation spectroscopy (FSC) were developed around the same time for single molecule detection and analysis in solution and were further optimized with applications in evolutionary biology.[28]

Approaches where the fluorescent product remains in the cell or is immobilized on the cell are compatible with high throughput FACS but suffer from potential cross-contamination and are incompatible with certain substrates. *In vitro* compartmentalization (IVC) in water-in-oil emulsions has emerged as an alternative to preserve the phenotype-genotype linkage.[24] IVC has attracted a lot of interest and has been developed in parallel to the advancement of research on directed evolution over the past 20 years. Surfactant-stabilized single- (water-in-oil) or double- (water-in-oil-in-water) emulsions constitute optimal compartments for directed evolution thanks to their long-term stability over a range of physicochemical factors including temperature, pH etc. Moreover, the formation of such compartments using microfluidic devices yields monodisperse droplets and allows for more controlled encapsulation of reactants.

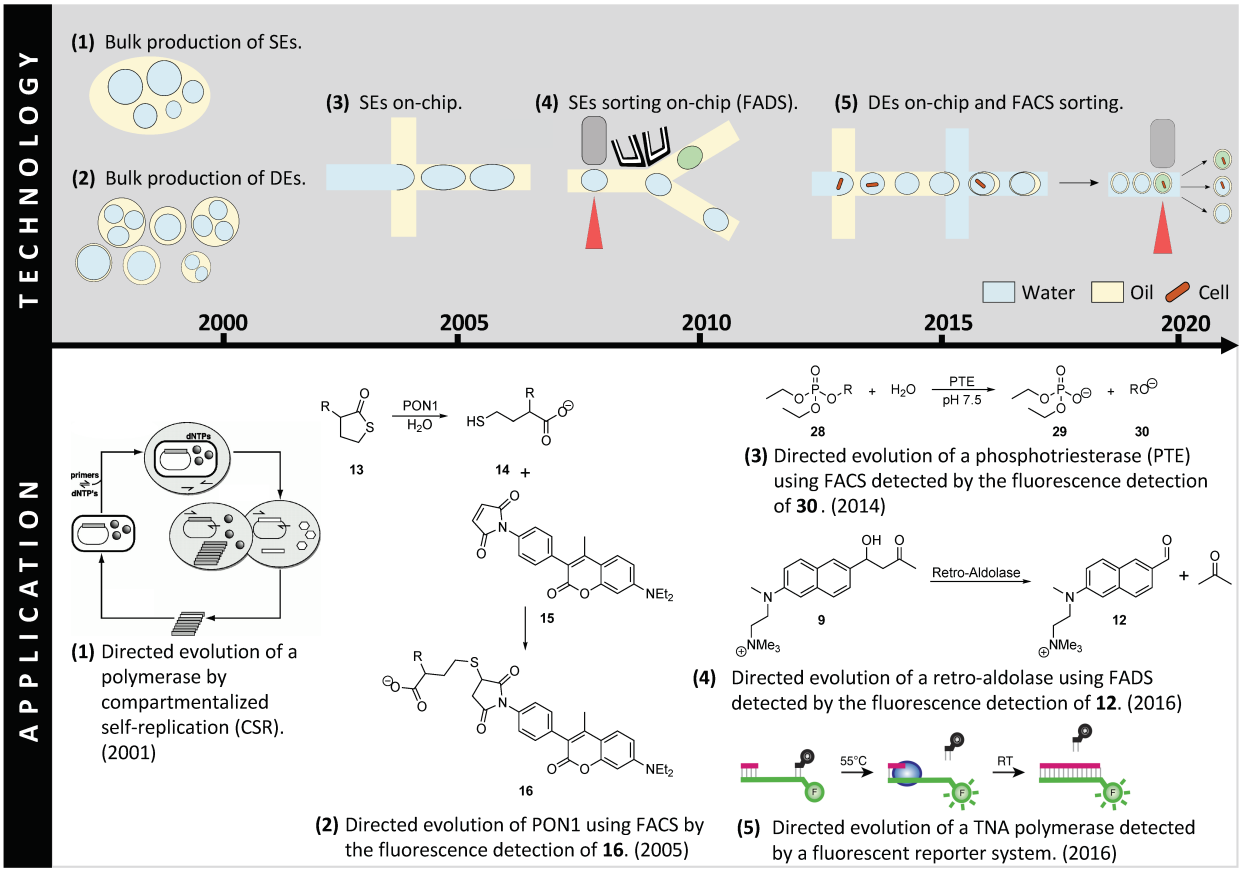


Figure 2. Milestones in the development of droplet microfluidics (top) and their applications to directed evolution (bottom) in the last twenty years. (1) Bulk production of single emulsions (SEs).[29] Directed evolution of a Taq DNA polymerase based on compartmentalized self-replication in SEs produced in bulk.[30] (2) Bulk production of double emulsions (DEs).[31] Directed evolution of *E. coli* surface-displayed serum paraoxonase 1 (PON1) using DEs produced in bulk.[32] (3) On-chip production of SEs.[33,34] Directed evolution of a phosphotriesterase through the encapsulation of E.coli expressing the enzyme on their surface in SEs produced on-chip. The SEs consist of a gellable liquid and form gel beads following a gelation step. The beads can be analyzed and sorted by FACS.[35] (4) On-chip sorting of SEs.[36,37] Directed evolution of a retro-aldolase using SEs formed on-chip and subsequent fluorescence-assisted droplet sorting (FADS) on-chip.[38] (5) On-chip formation of DEs followed by FACS sorting.[39,40] Directed evolution of a manganese-independent α-L-threofuranosyl nucleic acid (TNA) polymerase using DEs generated on-chip and subsequent sorting by FACS.[41] (Reprinted with permission (1) of the National Academy of Science USA. Copyright (2001) National Academy of Sciences. Reprinted (5) from reference[41]).

Directed evolution studies have directly benefitted from the development of droplet microfluidics, allowing faster screening of larger libraries. In turn, the need for more specific and powerful tools for directed evolution has driven research in droplet microfluidics forward. In the last twenty years, engineering and biochemistry research groups have worked together to improve existing systems and develop new ones (Figure 2). In this review, we provide a chronological overview of the intertwined development of microfluidics droplet-based compartmentalization methods and *in vivo* directed evolution of enzymes.

1.3 SINGLE EMULSIONS

Single emulsion droplets are aqueous compartments surrounded by an oil phase. The droplets can be stabilized using surfactants, i.e., amphiphilic molecules that arrange themselves at the water/oil interface. Methods of increasing complexity have been developed for the formation of such compartments, allowing improved control over the droplet size, the throughput or the reagents' encapsulation. Due to the external oil phase, water-in-oil (w/o) droplets are not compatible with commercially-available FACS devices. To overcome this challenge, methods for on-chip sorting have been implemented. Most recent devices have sorting throughputs of up to several kHz.[42]

1.3.1 TECHNOLOGY ADVANCES I: BULK EMULSIFICATION AND STRATEGIES FOR THE ENCAPSULATION AND IMMOBILIZATION OF REAGENTS AND REACTION PRODUCTS

Different methods are used for the production of w/o compartments. Bulk emulsification allows fast and simple formation of droplets, but has limited encapsulation efficiency and yields polydisperse droplets. Bulk emulsification techniques, such as stirring and emulsifier-based methods, were described before 1980.[29,43] Later studies focused on the characterization of the physical properties of emulsions produced with custom-made or commercially-available homogenizers,[44–46] highlighting that droplets of sizes ranging from 0.1 to 100 µm in diameter can be produced.

Whole cells or genetic material can be encapsulated in w/o droplets. The cell encapsulation follows the Poisson distribution and single cell compartmentalization can be achieved by adjusting the dilution of the cell-containing solution.[47] In the early 1990s, emulsions could be produced at high throughput but were incompatible with analytical tools with similar throughput. To circumvent this challenge, other droplet-based strategies were developed to screen active variants with FACS devices. One of the first techniques to emerge consisted in co-encapsulating an *in vitro* transcription and translation (ivTT) mixture with single microbeads, each displaying the gene encoding the protein of interest in w/o emulsions (Figure 3b).[48] In this study, antibodies bound to the streptavidin-coated microbeads could immobilize the translated proteins. Upon translation, the emulsions were ruptured to retrieve the microbeads and, subsequently, incubated with horseradish peroxidase (HRP) which bound to the proteins of interest via a ligand. In a second step, the beads were incubated with hydrogen peroxide and fluorescein tyramide, leading to the fluorescent labelling of the bead. FACS-sorting of the microbeads enabled the identification of a protein with high affinity towards the ligand used in the screen.

Another strategy enabling the use of FACS consists in generating droplets with a gellable liquid in which genes and either encoded enzymes or whole cells can be encapsulated. Through a cooling step, the droplets are converted into FACS-compatible gel beads, immobilizing and

compartmentalizing the genetic material (Figure 3c).[49] The relative permeability of gel beads favors the constant intake of growth medium or the addition of certain substrates at a later time point, while retaining the cell microcolonies. Using this technique, Weaver *et al.* encapsulated mammalian, bacterial and fungal single cells in agarose beads of 20-90 µm in diameter (Figure 3d).[50] After an incubation step in the growth medium and a staining step with fluorescent markers for biomass, the cell colonies were analyzed by FACS. In a related study, Sahar *et al.* analyzed the properties of the encapsulated bacterial colonies.[51] Among others, they characterized the intracellular esterase activity of a *P. aeruginosa* cell population. This was achieved through the addition of a fluorogenic substrate, 6-carboxy-fluorescein-diacetate, to the gel beads followed by an incubation step. They additionally described the activity of the secreted enzyme elastase by encapsulating its fluorescently-labelled substrate casein during droplet formation and determined the decrease in fluorescence caused by the leakage of the product out of the bead.

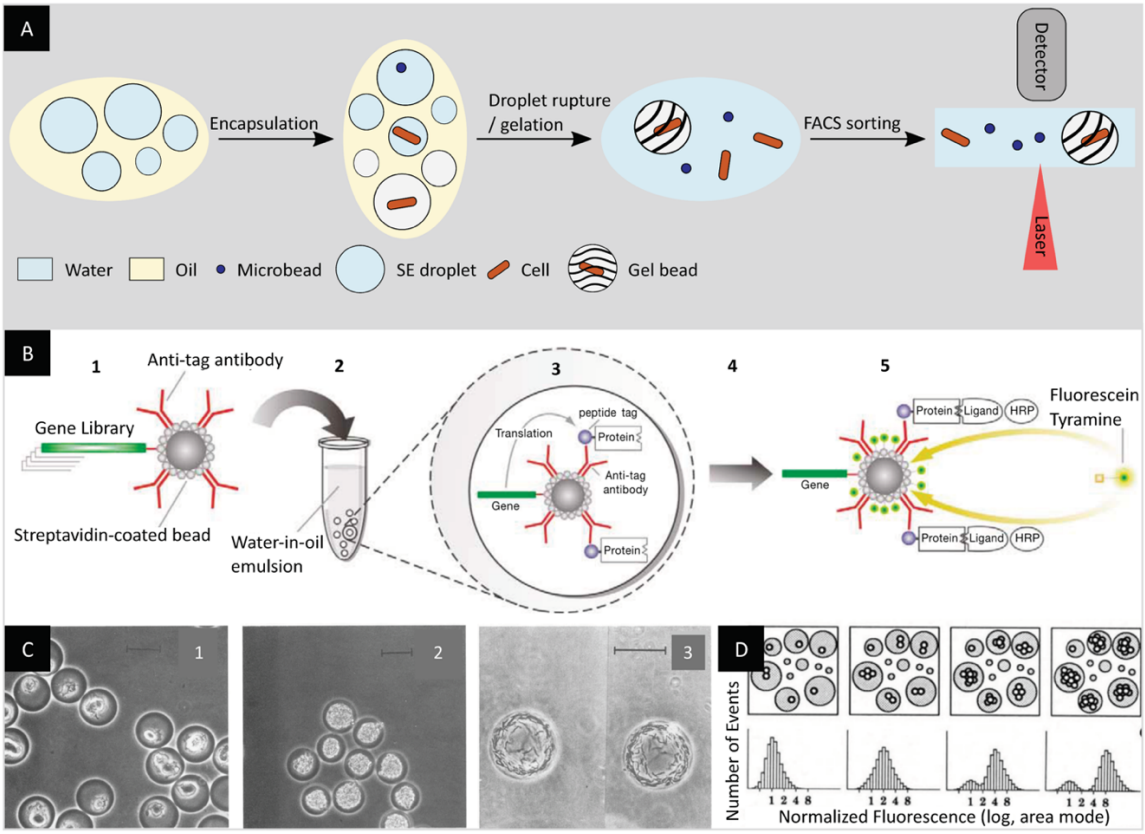


Figure 3. Bulk emulsification, encapsulation and sorting. a) Bulk emulsification permits the encapsulation in droplets of microbeads, cells or genetic material together with substrates and reagents. Once the reaction of interest has been carried out, the product can be immobilized on the microbeads, cells, or in the droplet itself following a gelation process. The gel beads can be directly analyzed and sorted by FACS, while the beads and cells require the droplet to be ruptured first. **b)** Co-encapsulation of an ivTT mixture with single DNA-coated microbeads in droplets for the identification of proteins with high binding affinity towards a ligand via FACS-sorting. (1) Streptavidin-coated microbeads displaying each a variant of a gene library encoding the protein of interest (2) are co-encapsulated with an ivTT mixture in w/o emulsions. (3) The translated proteins are immobilized by antibodies bound to the microbeads. (4) Droplet rupture permits the microbeads retrieval and (5) incubation with HRP, which binds to the proteins of interest. A second incubation step with hydrogen peroxide and fluorescein tyramide, labels the beads fluorescently and permits FACS-sorting for the identification of a protein with high affinity towards the ligand used in the screen.[48] **c)** Micrograph of the encapsulation and growth of different microbial cells in gel microbeads: *E. coli* (1), *S. cerevisiae* (2) and *M.*

xanthus (3). Scale bars: 20 µm.[49] **d)** Fluorescence-based biomass quantification of yeast cells trapped in gel microbeads.[50] (Images reprinted with the permission from (B) Wiley-WCH Verlag GmbH & Co KGaA, (C) the American Society for Microbiology, (D) Springer Nature Limited.)

1.3.1.1 APPLICATIONS I: *IN VITRO*

The first study using *in vitro* compartmentalization for applications in molecular evolution resulted from a collaboration between the Griffiths and Tawfik groups.[52] In this study, *in vitro* compartmentalization (IVC) of a single gene encoding either a DNA-methyltransferase HaeIII or a dihydrofolate reductase (DHFR) followed by ivTT led to the enrichment of an enzyme for DNA methylation (Figure 4). M.HaeIII genes encoding HaeIII and folA genes encoding DHFR, both containing a site designed for methylation/restriction by M.HaeIII, were encapsulated and tested for methylation efficiency. If M.HaeIII was present, the gene was methylated and was thus not digested in the subsequent digestion step with the endonuclease HaeIII. On the other hand, if DHFR was present, the gene was not methylated and was therefore digested by the HaeIII endonuclease.

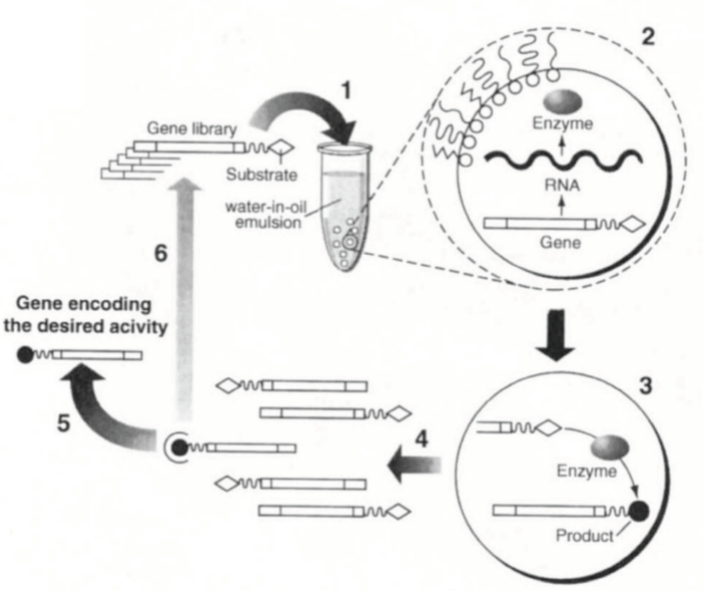
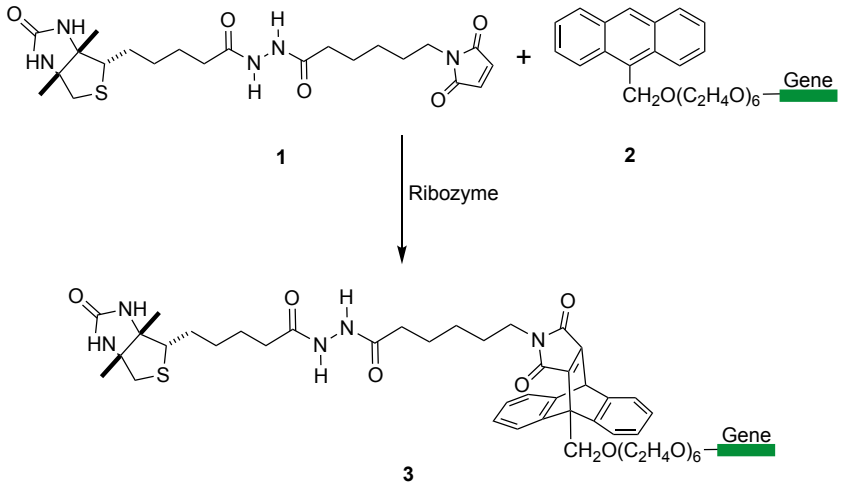


Figure 4. Schematic representation of an IVC selection strategy for DNA/RNA. (1) One single gene linked to a substrate is compartmentalized within a w/o emulsion. (2) ivTT yields a functional protein/RNA and (3) an enzymatic reaction converts a substrate into a product which remains linked to the gene. (4) After the emulsions are ruptured, (5) the genes linked to the product are selectively enriched and (6) either characterized and/or encapsulated for another round of evolution.[52] (Reprinted with permission from Springer Nature Limited.)

Methylated HaeIII sites resistant to digestion were amplified using PCR and analyzed on an agarose gel. Model enrichment of a library starting with 0.1% M.HaeIII led to a 500-fold enrichment in a single cycle. The same approach was used in a follow-up study to improve the sequence specificity. A more active species was selected from a random mutagenesis library at three positions with $\sim 3.3 \cdot 10^7$ variants. Remarkably, over only two rounds of screening, 11 variants with up to ~19-fold improvement were identified. All identified hits bore two mutations, whereas the third position

revealed itself to be crucial for the methyltransferase activity and did not tolerate any other mutation.[53]

With the aim of bringing the technology to the next level, single emulsions were used to evolve ribozymes for a bimolecular Diels-Alder reaction. In a larger evolution campaign consisting of four rounds of IVC, ribozymes catalyzing the intermolecular Diels-Alder reaction between 9-anthracenylmethyl hexaethylene glycol (AHEG, **1**) and biotin-maleimide (**2**) with multiple-turnovers were evolved (Scheme 1). After four rounds of evolution, variants with a catalytic efficiency $k_{cat}/(K_{m1}\ K_{m2})=5.3\ 10^5\ M^{-2}\ s^{-1}$ were identified. These artificial enzymes display efficiencies that are comparable to catalytic Diels-Alderase antibodies.[54] Using a custom-built homogenizer, Paegel and Joyce evolved RNA enzymes with ligase activity, selecting enzymes that could resist inhibition by neomycin. A library of 10^{11} variants was evolved over 5 rounds to obtain mutants with better tolerance to neomycin and generally higher K_m values.[55]



Scheme 1. The intermolecular Diels–Alder reaction between biotin-maleimide (**1**) and AHEG (**2**). The gene to be evolved is covalently attached to **2** and **3**, but only active catalysts give products **3**. The Diels-Alder products **3** are subsequently captured by streptavidin-coated magnetic beads allowing their downstream PCR enrichment.[56]

1.3.1.2 Applications II: *in vivo*

Meanwhile, the first *in vivo* applications using single emulsions were reported. The screening approach used for the first studies was based on compartmentalized self-replication (CSR) (**Figure 5**a). The directed evolution of Taq DNA polymerase was carried out in polydisperse emulsions generated by stirring. With this approach, Ghadessy *et al.* identified a Taq DNA polymerase variant with elevenfold increased thermostability and a variant with over 130-fold increased resistance to the inhibitor heparin (**Figure 5**b).[30] Other similar approaches of CSR in single emulsions involved the directed evolution of the same Taq polymerase for broader substrate scope and faster-cycling mutants (35-90-fold higher affinity for the primer, twofold increase in extension rate).[57] To expand the technology to non-polymerase type enzymes, cooperative CSR was applied to evolve a nucleoside diphosphate kinase (NDK). NDK converted dNDPs to dNTPs which, in turn, could be

used by a polymerase to replicate the genetic material. In this manner, only genes encoding active NDK were replicated thus affording a straightforward approach to evolve simple cascade reactions.[30]

The systems described above rely mostly on self-replication. Retaining a fluorescent signal on the encapsulated species itself is an essential feature to allow for FACS sorting. This was illustrated with yeast cells encapsulated in droplets. A library of yeast cells with surface-displayed glucose oxidase (GOx) and horse radish peroxidase (HRP) was encapsulated and screened for the conversion of glucose to gluconolactone (Figure 6). The hydrogen peroxide byproduct of this reaction was reduced by HRP leading to the generation of a fluorescein tyramide radical which, in turn, reacted with a tyrosine residue on the surface of the yeast cell. In this manner, the yeast cells retained the fluorescent information and could be sorted after rupturing the emulsions. From a library containing 10^5 variants, resulting from error-prone Polymerase Chain Reactions (epPCR), a variant with five mutations and a 2.7-fold improvement in k_{cat} was identified.[58] Similarly, GOx was evolved for different conditions, resulting in twofold improved thermal stability compared to wild type ($t_{1/2}$ ~20 min at 60°C) as well as a fourfold and 5.8-fold improvement in k_{cat} at pH 5.5 and pH 7.4 respectively.[59] Recently, GOx was coupled to the yeast-enhanced green fluorescent protein (yGFP) to afford a chimera allowing the simultaneous detection of the protein expression level and the activity of the same enzyme. This system led to a 2.5-fold enrichment of expressed, active variants and a 2.3-fold increase in V_{max} in just one round of screening.[60]

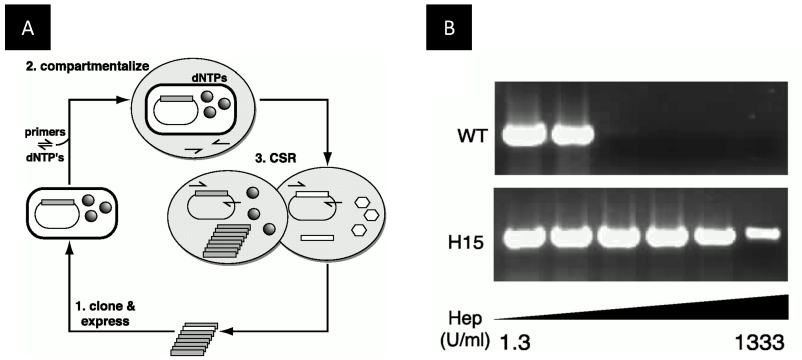


Figure 5. First *in vivo* directed evolution study in single emulsions based on CSR. a) Compartmentalized self-replication (CSR): (1) Genes encoding a polymerase are cloned and expressed in *E. coli* and encapsulated in w/o droplets together with primers and dNTPs (2). Poorly active variants (hexagons) cannot replicate efficiently, whereas functional and active variants of the polymerase enzyme (spheres) result in self-replication (3) and can be extracted and analyzed or recloned for another cycle of CSR. **b)** PCR-activity of the wild type and evolved mutant (H15) in the presence of heparin.[30] (Reprinted with permission from National Academy of Science USA. Copyright (2001) National Academy of Sciences.)

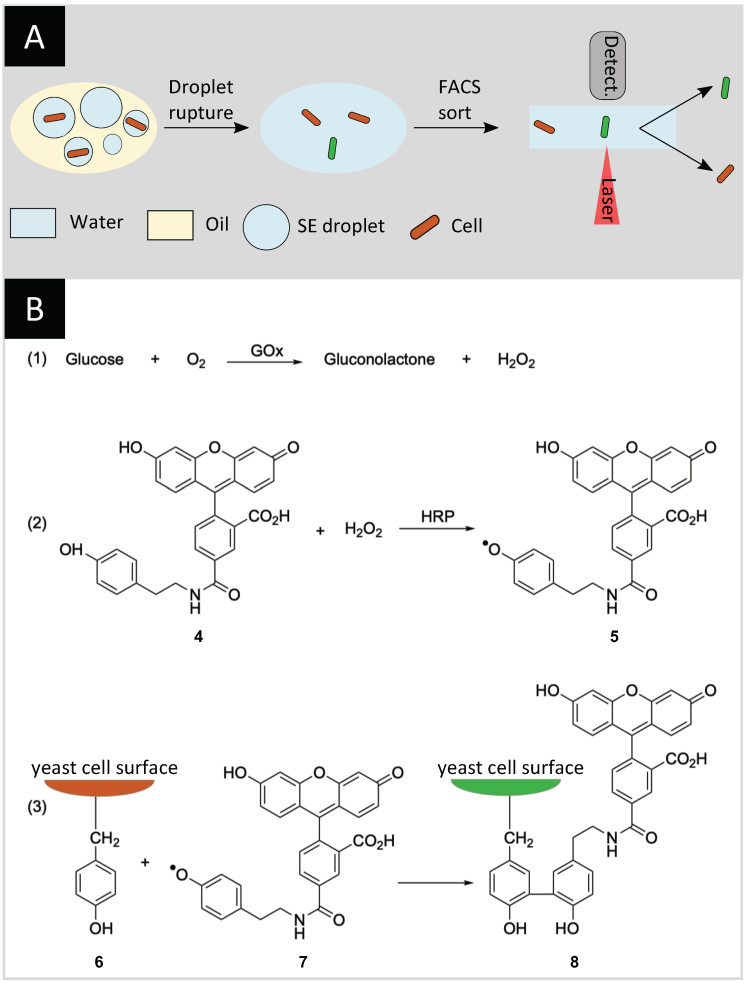


Figure 6. Screening of yeast surface-displayed GOx. a) Yeast cells with surface-displayed GOx and HRP are encapsulated in w/o emulsions. The catalytic reaction leads to stained yeast cells which, after rupturing the emulsions, are amenable to FACS sorting. **b)** GOx converts glucose into gluconolactone (1) and the byproduct H_2O_2 is reduced by the HRP to produce a fluorescein tyramide radical **5** (2). The radical then reacts with a tyrosine residue **6** on the surface of the yeast cell, leading to a stained yeast cell **8** (3).[58]

1.3.1.3 Applications III: Encapsulated Microbeads

A prominent example involving microbeads consists in the directed evolution of an extremely efficient phosphotriesterase (PTE) using streptavidin-coated microbeads.[35] In this study, Griffiths and Tawfik used polystyrene microbeads displaying single genes anchored via a biotin-streptavidin linkage. Within the w/o emulsions, multiple copies of PTE were produced by ivTT and anchored to the bead using an antibody. The emulsions were then ruptured and the beads were re-encapsulated to add a soluble biotin-tagged substrate. The catalysis was performed inside the emulsions and the biotin-tagged product was retained on the bead. Subsequent rupture of the emulsions and labelling with a fluorescent anti-product antibody, facilitated the sorting of active species. Relying on this approach, the authors identified a variant with a k_{cat} of 10^5 s^{-1} after six rounds of directed evolution from a $3.4 \cdot 10^7$ library. This corresponds to a 63-fold improvement over the wild type enzyme. This work paved the way for similar approaches such as the enrichment of an oxygen-tolerant [Fe-Fe] hydrogenase I from *C. pasteurianum* (CpI) for the reduction of the fluorogenic compound C_{12}-

resazurin[61] and the mock enrichment of a wild type HRP immobilized on microbeads.[62] Notable directed evolution efforts include the directed evolution of *i)* a *trans*-acting Bartel class I ligase with up to 90-fold rate enhancement[63] and *ii)* a sortase A from *Staphylococcus aureus* with a 114-fold enhancement in k_{cat}/K_m.[64] Recently, the Panke and co-workers used alginate beads for the directed evolution of the broad-spectrum amino acid racemase from *Pichia pastoris* (*Pp*AAR) for the racemization of *D*-ornithine, an interesting target for industrial applications. Starting from a library with $1.2 \cdot 10^7$ variants, they observed an up to 2.7-fold k_{cat}/K_M improvement over wild type after three rounds of directed evolution.[65]

1.3.2 Technology advances II: Microfluidics-based droplet formation and reduction of cross-talk between droplets

The need for higher droplet monodispersity and better control over the formation process led to the development of the first microfluidic chips for droplet production.[66,67] Compared to bulk emulsion methods, microfluidics-based methods require the fabrication of the device and its operation, but allow for high-throughput encapsulation in monodispersed compartments. The first study by Thorsen *et al.* introduced a device with a T-shaped junction to generate emulsions at a frequency of 20-80 Hz. (**Figure 7**b).[33] Similarly, Anna *et al.* displayed the controlled formation of droplets in a flow-focusing channel geometry for the production of emulsions with diameters as small as 10 µm (**Figure 7**c).[34] Similar channel geometries were used later to analyze reagent mixing inside droplets, spontaneous merging of droplets of different sizes, reagent addition to droplets and droplet splitting.[68,69]

For directed evolution, each droplet ideally contains one cell. However, cell encapsulation using microfluidic devices follows the Poisson distribution, resulting in a majority of empty droplets at low cell concentrations, thus lowering the effective throughput. Yet, unlike bulk methods, several studies have highlighted the possibility of overcoming Poisson's distribution limitations on a microfluidic device. Using particular channel geometries and hydrodynamic effects at high flow rates to order cells, various groups succeeded in yielding up to 80% single-cell-containing droplets (**Figure 7**d).[70,71]

Another critical aspect of droplet microfluidics compatibility with directed evolution resides in the ability of the droplets to retain the substrate and product of the enzymatic reaction of interest. In their study, Courtois *et al.* investigated the leakage of fluorescein-based substrates from droplets into the oil and succeeded in improving the retention to more than 18 hours by addition of bovine serum albumin (BSA). The authors illustrated the versatility of their system by characterizing the enzymatic activity of alkaline phosphatase expressed by *E. coli* cells and distinguishing empty droplets from cell-containing droplets.[72] The retention of substrate and product in w/o emulsions

can also be addressed with the use of gel beads as described earlier, which can be created as well by droplet microfluidics (**Figure 7**e).[73]

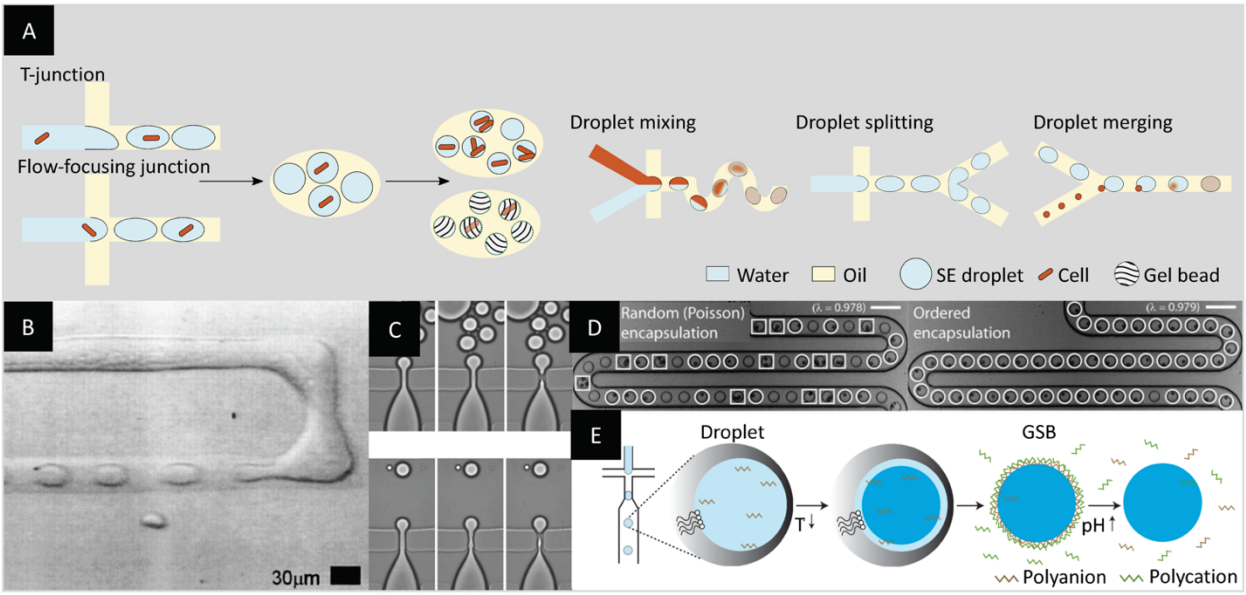


Figure 7. Microfluidics-based droplet formation of monodispersed w/o emulsion droplets. a) The use of microfluidic chips with channels forming either a T-junction or a flow-focusing junction allows the formation of highly monodisperse droplets at high throughput. Specific channel geometries can be used to improve droplet mixing, splitting or merging. **b)** Micrograph of the first T-junction design for droplet production at 20-80 Hz.[33] **c)** Micrograph of the first PDMS chip with flow-focusing junction for the formation of droplets as small as 10 µm in diameter.[34] **d)** Micrograph of the comparison between random bead encapsulation in droplets with improved bead encapsulation using cell alignment resulting from hydrodynamic interactions. Droplets containing two particles or more are highlighted by a square while droplets encapsulating single particles are highlighted by a circle. Scale bars: 100 µm.[70] **e)** Microfluidic production of droplets with a gellable liquid for the formation of gel beads with a semipermeable polyelectrolyte shell (gel-shell beads, GSBs). The shell can be ruptured under basic conditions.[73] (Images reprinted with permission from **b)** American Physical Society, **c)** AIP Publishing LLC, **d)** The Royal Society of Chemistry, **e)** Springer Nature Limited.)

1.3.2.1 APPLICATIONS I: *IN VITRO* AND *IN VIVO*

As early as 2013, Scanlon *et al.* presented the application of hydrogel emulsions produced on chip for the discovery of natural product-based antibiotics. A recombinant antibiotic-producing microbe (*Saccharomyces cerevisiae* or *E. coli*) was co-encapsulated with the pathogen (*S. aureus*) and a fluorescent label for dead cells.[74] After incubation, the emulsions were ruptured and the cells were sorted using FACS, allowing the identification of yeast or bacteria with bactericidal properties. In a model sort with a ratio 1:10'000 of positive control yeast (secreting the bacteriolytic enzyme lysostaphin and constitutively expressing yEGFP) and negative control yeast (ineffective against *S. aureus,* non-fluorescent), the authors reported a complete enrichment over three sorting rounds. Applying a similar methodology, *E. coli* encapsulated in hydrogel emulsions were screened for pBAD promoter activity.[75] In this study, a library of single *E. coli* cells expressing GFP were encapsulated in hydrogel emulsions using a microfluidic device. Depending on the promoter sequence, differences in GFP expression allowed FACS sorting based on GFP fluorescence. After sorting and enzymatic digestion of the agarose, the microcolonies were plated on agar plates and analyzed to find an averaged 1.25-fold improvement in protein expression in one round of screening.

In a recent study, Fischlechner *et al.* described the formation of gel beads with a polyelectrolyte shell. This shell led to the retention of significantly smaller molecules, with a molecular weight cut-off 200-fold lower than gel beads reported previously. They reported the co-encapsulation of single *E. coli* cells expressing a variant of a phosphotriesterase and a fluorogenic substrate. Subsequent lysis of the *E. coli* in the gel beads released the active enzyme catalyzing the hydrolysis of a phosphotriester to yield a fluorescent product. The beads retained the fluorescent product and the beads containing the most active variants were selected using FACS at rates $>10^7$ Hz. A variant with almost twentyfold higher k_{cat}/K_M could be identified in a single round.[73] Similar approaches have been used to evolve production hosts for industrially-relevant enzymes: improved overexpression and 1.3-fold higher secreted amounts of xylanase by *P. pastoris* in gel microdroplets was recently reported by Ma *et al.*[76]

1.3.3 TECHNOLOGY ADVANCES III: ON-CHIP OBSERVATION, MANIPULATION AND SORTING OF DROPLETS

Most of the technological developments described so far were optimized for on one-step processes and reactions. However, many reactions require multiple steps where the addition of new reagents is required or the different reaction conditions are not compatible with each other. With the development of droplet-based microfluidics more options for generation, fusion, control and analysis of droplets have emerged (Figure 8a).[68,77]

An important advancement of droplet-producing microfluidic devices consists in the integration of electrodes generating an electric field across microfluidic channels. In an early study, Chabert *et al.* investigated the electrocoalescence of w/o pair droplets as a tool for reagent addition. Using AC fields, they succeeded in displacing and merging droplets of 600 µm in diameter under static and flow conditions.[78] Another study established the high-throughput electrocoalescence of pair droplets in a PDMS microfluidic chip (Figure 8b).[79] Two flow streams with droplets of different diameters (13-50 µm) merged into a single channel with downstream electrodes generating an electric field. As the droplet velocity is size-dependent, the size mismatch allowed the smaller and larger droplets to form pairs upon contact and led to the subsequent electrocoalescence as the pair passed through the electric field. The method was illustrated by determining the k_{cat} of an enzymatic reaction through the encapsulation of β-galactosidase and its fluorogenic substrate, resorufin-β-D-galactopyranoside, in pair droplets. Pioneering studies led to the development of controlled reagent injection to droplets with higher throughput. Abate *et al.* proposed the use of picoinjectors to add reagents to droplets at frequencies of several thousand Hertz (Figure 8c).[80] Recent developments for precise reagent delivery inside w/o droplets still involve electric fields and

more complex systems such as the rupture of triple emulsions[81] or the use of a 3-phase flow.[82] Yet, the use of electrodes is not restricted to droplet merging but also allows the displacement of droplets.[37,83] In an innovative study, Ahn, Kerbage *et al.* reported on a droplet sorter based on the use of dielectrophoretic forces to direct droplets towards either side of a microfluidic junction.[84]

In parallel to the development of droplet manipulation, several groups focused on on-chip fluorescence detection for reaction monitoring in droplets.[85] In an early study, Dittrich *et al.* encapsulated ivTT mixture and red-shifted GFP-encoding (rsGFP) genes in w/o droplets and monitored the protein expression on-chip using fluorescence spectroscopy.[86] The combination of sensitive fluorescence detection and dielectrophoretic sorting led to the development of the first fluorescence-activated droplet sorting (FADS) device.[36] In this joint effort of the Weitz and Griffiths groups, *E. coli* cells expressing either *β*-galactosidase or an inactive variant were co-encapsulated with a fluorogenic substrate. The groups sorted the droplets based on enzymatic activity at a rate of 300 Hz with a low error rate. In a later study by the same authors, a system with three chips for droplet production, reagent addition *via* pair droplet fusion and fluorescence intensity-based sorting, was used for the kinetic monitoring of *in vitro* translated laccase.[87] Decoupling the processes allowed the handling of droplets at different rates, 7000 Hz for droplet production and 3000 Hz for droplet merging.

A similar decoupled process was used to enrich an active variant of *in vitro* translated *β*-galactosidase.[90] There, the droplets containing active variants were merged on-chip with an aqueous stream for easier retrieval of the genes. Similarly, Svahn and coworkers reported on the enrichment of a yeast strain based on its enzyme production using FADS. They achieved an enrichment close to the theoretical maximum and identified a clone with twofold increase in amylase production over a single round of screening (Figure 8d).[88] In a similar study, Ostafe *et al*. used FADS to enrich cellulase-producing yeasts from an inactive cell population with an enrichment factor of up to 300-fold.[91] Recent improvements of FADS devices involve multi-way sorting: Frenzel *et al.* proposed a chip allowing for droplet sorting in four outlets at a maximal throughput of 2-3 Hz.[92] A later study by Caen *et al.* described a five-way sorting system with an almost 100-fold higher throughput (Figure 8e).[89] Recently, a microfluidic chip for the sorting of w/o droplets based on fluorescence life-time was reported by Hasan *et al*.[93] Most of the examples described above for dielectrophoresis-based sorting rely on the use of fluorogenic substrates. However, recent studies propose label-free techniques for on-chip sorting. Alternatives such as interfacial tension-based sorting to distinguish between live and dead cells[94] or intelligent image-based sorting capable of analyzing images and taking sorting decisions in real-time[95,96] offer interesting prospects.

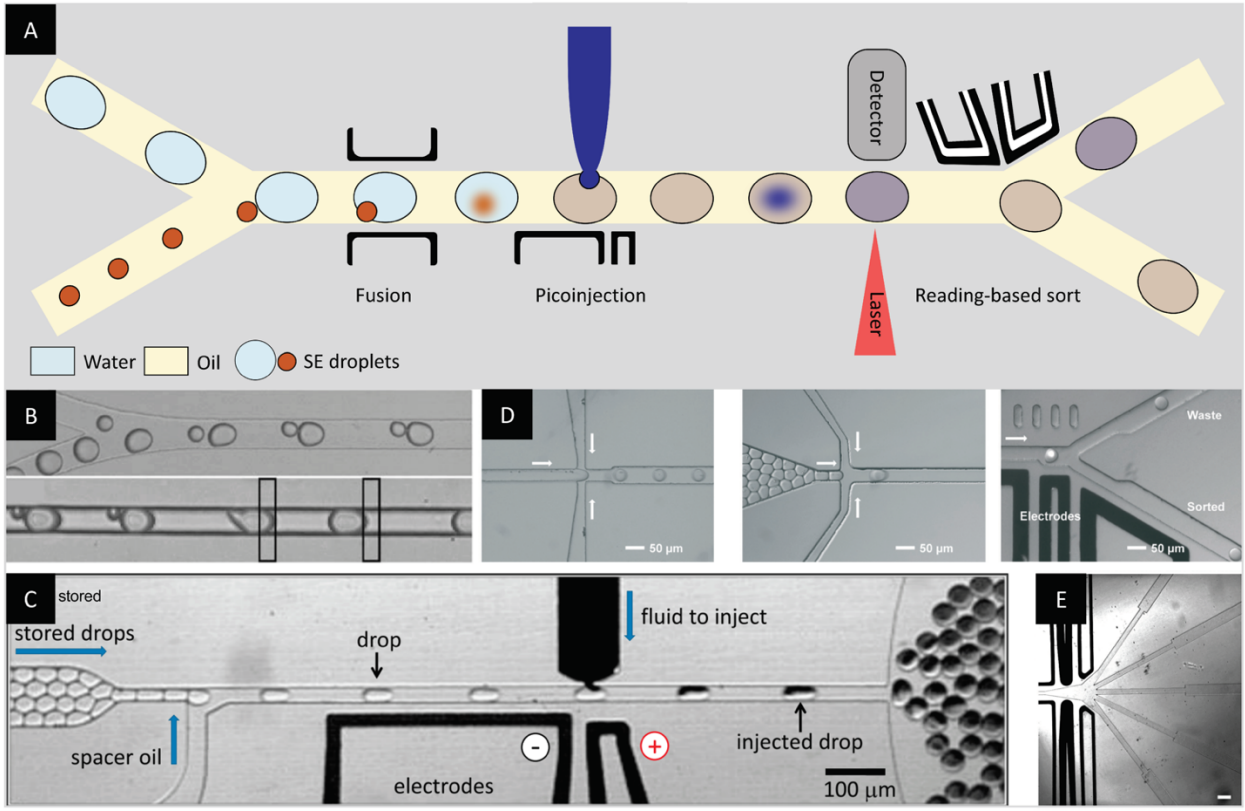


Figure 8. Manipulation, observation and sorting of droplets on-chip. a) Integrated electrodes allow the controlled addition of a reagent to pre-formed droplets at a later time-point, either by fusing droplets (orange) or by picoinjection (blue). Droplet fluorescence or absorption can be monitored on-chip and another set of integrated electrodes can sort the droplets in different channels based on the reading. **b)** Micrograph of the electrocoalescence of droplet pairs.[79] **c)** Micrograph of the addition of reagent to pre-formed droplets using picoinjection.[80] **d)** Micrographs of droplet production on-chip, followed by the droplet reinjection in a second chip, using oil as spacer, and the droplet sorting by dielectrophoresis based on fluorescence detection.[88] **e)** Micrograph of a sorting junction designed for 5-ways sorting of droplets. Scale bar: 200 µm.[89] (Images reprinted with permission from **b)** AIP Publishing LLC, **c)** National Academy of Science USA, **d)** The Royal Society of Chemistry. Image reprinted **e)** from reference [89]).

1.3.3.1 Applications I: *in vitro*

An initial effort involving a multi-step process was highlighted using the example of a CotA sporulation protein, a laccase from Bacillus subtilis catalyzing the oxidation of various aromatic compounds using molecular oxygen as oxidant. Since the laccase assay was incompatible with the *in vitro* protein expression system, sequential addition of reagents at different time points were required.[87] Following this strategy, Fallah-Araghi *et al.* developed a completely *in vitro* platform for the screening of active lacZ genes encoding the enzyme *β*-galactosidase starting from single genes. In their study, genes were encapsulated before on-chip electro-coalescence and sorting. In a model sort of active lacZ genes *vs.* inactive lacZmut with a ratio of 1:100, they reached a 502-fold enrichment in a single round of screening.[90] Building on these examples, Goto *et al.* reported a device for the encapsulation and sorting of nanoliter droplets and applied the method on the model screening of an isocitrate dehydrogenase (IDH) from *Streptococcus mutans*. Starting from a library

of 103 variants and two rounds of screening, they isolated a variant with about threefold higher activity than wild type.[97]

1.3.3.2 APPLICATIONS II: *IN VIVO*

Technological progress led to up to 1000-fold faster screening and a million-fold decrease in reagent costs as exemplified by the joint efforts of Abate, Baret, Griffiths and Weitz. In their seminal study, they applied on-chip droplet generation and dielectrophoretic sorting for the high-throughput screening of HRP displayed on the surface of yeast cells. After sorting, the droplets were ruptured, thus making the most active yeast cells readily available for the next round of mutagenesis and sorting. With this screening platform, they screened libraries with up to 10^7 variants and achieved an overall ~sevenfold improvement in catalytic efficiency over 9 rounds. The highest catalytic efficiency reached ~$2.5 \cdot 10^7$ M^{-1} s^{-1}, thus approaching diffusion-limited efficiency (i.e. 10^8 M^{-1} s^{-1}).[98] Similar set-ups involving yeast encapsulated in single emulsions were used e.g. for the evolution of thermostable xylanase with improved activities (up to 4.7-fold)[99] or the improvement of yeast cells as production hosts (2-fold increase in α-amylase production).[88]

One of the first examples expanding the repertoire to *E. coli* was reported in 2015 by Abate and coworkers. They mapped protein sequence-function relationships by combining microfluidics with next generation sequencing and analyzing both sorted and unsorted populations. Starting from a library of $6 \cdot 10^7$, they enhanced glycosidase activity at higher temperatures in a single round of mutagenesis. The deep mutational scanning revealed regions which might be crucial for glycosidase activity, but also highlighted known patterns with mutational tolerance which were in accordance with examples from the enzyme family.[100]

Directed evolution is especially versatile if the initial catalytic activities are low, as for example for *de novo* designed biocatalysts. In another study, Hilvert and coworkers applied a microfluidics-based screening coupled to FADS to evolve a retro-aldolase by amine catalysis.[101] A previously designed retro-aldolase capable of cleaving a carbon-carbon bond in a non-natural substrate ((±)-methodol) *via* an enzyme-bound Schiff-base intermediate (**11**) showed modest catalytic efficiency (k_{cat}/K_M = 0.19 $M^{-1} \cdot s^{-1}$) and enantioselectivity (ee = 33% for (*S*)-methodol).[13] It was used as the starting point for the directed evolution campaign and the authors were able to significantly improve the catalytic activity. *E. coli* expressing the protein of interest in the cytoplasm were encapsulated in w/o emulsions with lysis buffer to release the enzyme in the droplet and a fluorogenic, charged methodol derivative (**9**) (Figure 9). Six focused libraries with up to five simultaneously mutated residues were screened and a variant with almost 80-fold increase in k_{cat} was identified in a single round of screening. Strikingly, the same mutations were identified in an earlier study on the same enzyme. This previous study, relying on a medium-throughput screening

campaign using MTPs, required five rounds of directed evolution to install these mutations. The best performing mutant of the microfluidics-based study was identified after only two rounds of evolution. It included ten mutations and exhibited a 73-fold increase in k_{cat}/K_M and tenfold preference for (*S*)-methodol.[101] The kinetics of the catalytic system were further optimized to give $>10^9$ rate enhancement, thus approaching Class I aldolase activities (natural enzymes catalyzing reversible carbon-carbon bond forming reactions) and accommodating a wider substrate scope.[102] The same group reported the isolation of an active cyclohexylamine oxidase (CHAO) identified from a single screening round of a library with 10^7 variants. They remodeled the active site of CHAO, achieving up to 960-fold increase in catalytic efficiency thus approaching the wild type levels of activity for a non-natural substrate.[103]

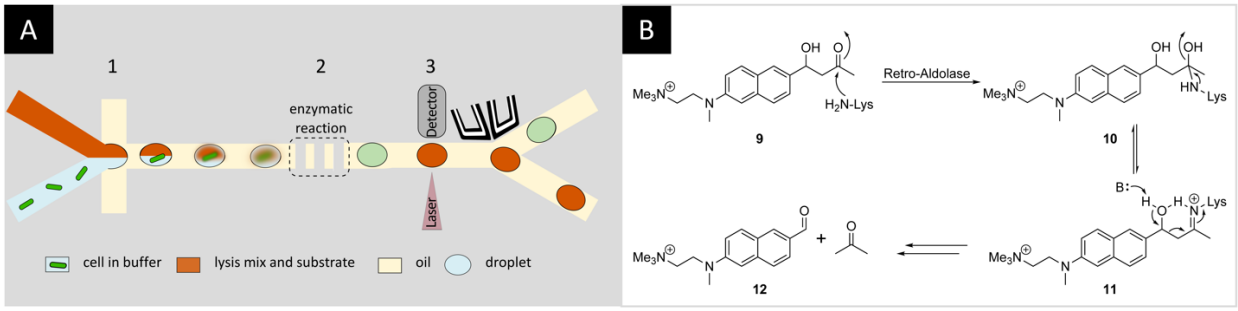


Figure 9. Directed evolution of a retro-aldolase using microfluidics-based FADS. a) Microfluidics-based FADS. *E. coli* are encapsulated with a substrate/lysis mix in w/o droplets (1). The cells are lysed in the droplets making the expressed retro-aldolase (orange) readily available and converting the aldol substrate (red) into a fluorescent product (green) (2). Finally, the droplets are sorted on-chip by activating the sorting electrodes when the fluorescence signal exceeds a certain threshold (3). **b)** Retro-aldolase-catalyzed cleavage of a charged methodol derivative (**9**) *via* an enzyme-bound Schiff-base intermediate (**11**) yields a fluorescent naphthaldehyde derivative (**12**) and acetone. The positive charge on the substrate/product ensures their retention in the droplets.[38]

1.4 Double Emulsions

Water-in-oil-in-water (w/o/w) double emulsions can overcome some of the remaining challenges with single emulsions, such as limited stability (e.g., shrinkage of droplets) and the need for rapid on-chip analytical methods. Although the creation requires an additional emulsification step,[104] the handling of w/o/w double emulsions on- and off-chip offers intriguing advantages, particularly, the compatibility with commercially-available FACS sorting devices is noteworthy.[105,106]

1.4.1 Technology Advances I: Bulk Emulsification and Development of Compatible Assays

Much like for single emulsions, the first methods developed for the formation of double emulsions were based on bulk emulsification using stirring or emulsifiers.[107] Double emulsions formed using these strategies are polydisperse and the inner aqueous phase compartment is usually made out of multiple compartments (Figure 10b). In an early study, Ficheux *et al.* scrutinized the

stability *vs.* coalescence of the inner aqueous compartments of double emulsions. Their research highlighted that the double emulsion's stability is mostly affected by the surfactant type (water- or oil-soluble) and concentration, and can vary on a scale that ranges from minutes to months.[31] Following these findings, other research groups investigated the stability of double emulsions with different surfactants and composition of outer aqueous phases.[108,109] A two-step emulsification process using a Couette mixer for the formation of quasi-monodisperse double emulsions with multiple aqueous compartments was proposed by Goubault and coworkers.[110]

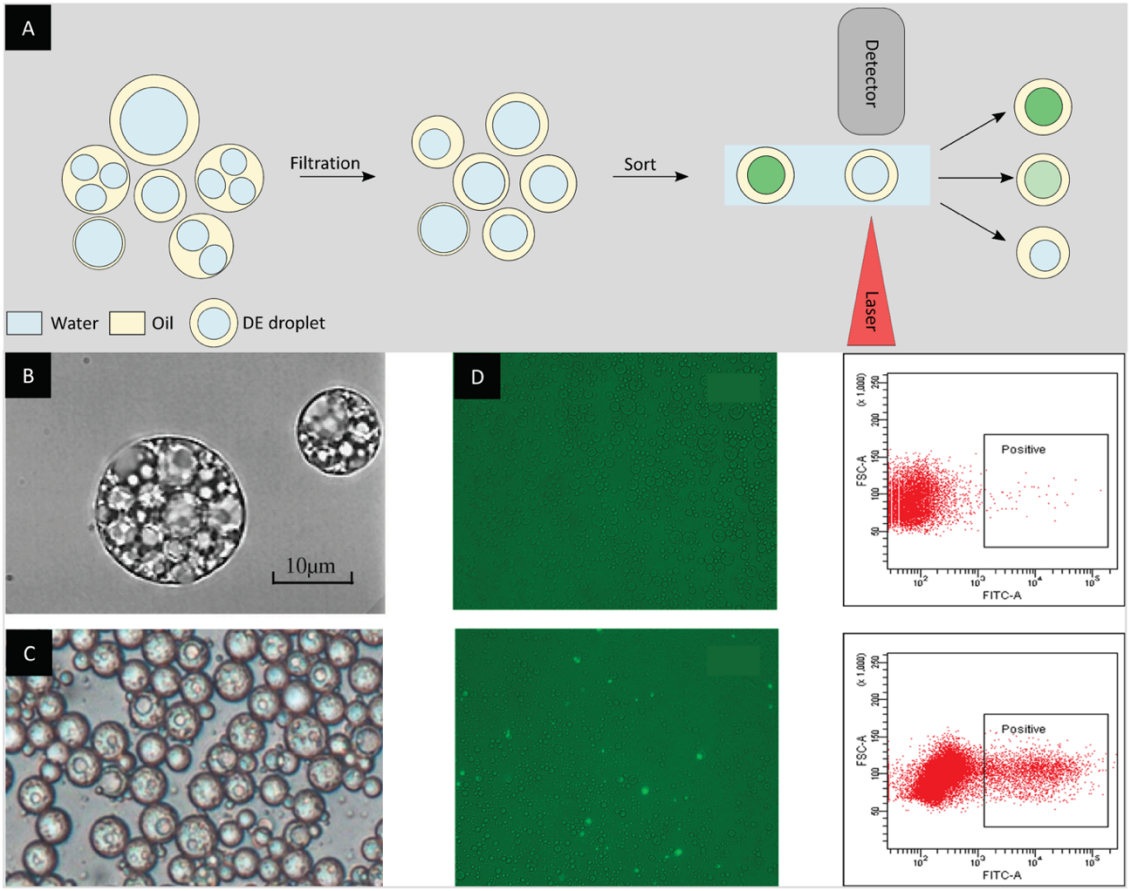


Figure 10. Bulk emulsification for the formation of double emulsion droplets. a) Double emulsion droplets formed by bulk emulsification are polydisperse and can contain multiple inner aqueous phase compartments. An additional filtration step can improve the sample homogeneity and lead to more reliable FACS sorting. **b)** Micrograph of double emulsion droplets resulting from bulk emulsification.[109] **c)** Micrograph of double emulsion droplets after membrane-extrusion.[112] **d)** Micrograph of double emulsions encapsulating either *E.coli* cells containing a plasmid for the expression of esterase on the cell surface or negative *E.coli* cells. The use of a fluorescein-based fluorogenic substrate allows the identification and sorting of double emulsions with cells displaying esterase activity as shown in the FACS plot.[112] (Image reprinted with permission from **b)** Elsevier. Images reprinted from **c)** and **d)** reference [112].)

More recent studies involving double emulsions generated in bulk focus on improving the compatibility of double emulsions with different screening methods. In their study, Prodanovic *et al.* proposed a fluorescent cascade assay in double emulsions for sorting enzyme libraries by FACS. The assay allowed the screening of a glucose oxidase gene library with 10^4 mutants based on the hydrogen peroxide production with a 50-200-fold enrichment factor.[111] To overcome the limitations imposed by the polydispersity of bulk double emulsions, Ma *et al.* improved the

production method by using membrane-extrusion, leading to the generation of more uniform double emulsions. The advantages of this method were illustrated by enriching a population of *E. coli* cells with esterase activity more than 300-fold (Figure 10c/d). The method was further applied to the directed evolution of a thermophilic esterase AFEST, resulting in a twofold improvement in catalytic activity as well as the identification of several mutants with k_{cat}/K_m values approaching diffusion-limited efficiency. [112]

1.4.1.1 APPLICATIONS: *IN VITRO* AND *IN VIVO*

The first application of double emulsions to directed evolution involved the model enrichment of the above mentioned FolA/M.HaeIII system. Positive w/o/w droplets containing FolA and the fluorescence marker FITC-BSA and negative w/o/w droplets containing M.HaeIII and BSA were produced separately and mixed in different ratios before sorting with a commercial FACS device. The positive and negative droplets were mixed in a 1:100 ratio and within one round of sorting a ~30-fold enrichment was observed.[113] Using the same technique, Griffiths and coworkers reported the first completely *in vitro* directed evolution campaign using double emulsions for the evolution of Ebg, a protein of unknown function. Starting with negligible activity, they screened a library of 2·106 members over four rounds of directed evolution and identified variants with *β*-galactosidase activity with at least 300-fold higher k_{cat}/K_M values compared to wild type Ebg.[114] The first study involving *in vivo* directed evolution in double emulsions was based on *E. coli* surface-displayed serum paraoxonase 1 (PON1). PON1 is a mammalian enzyme capable of hydrolyzing a broad range of substrates, in particular the homocysteine thiolactone, and thereby eliminating toxic metabolites. In a two-step process using a homogenizer, single emulsions containing *E. coli* with surface displayed PON1 were produced. The substrate (**13**) and a thiol-detecting dye (**15**) were then added via the oil-phase and a subsequent emulsification step led to the generation of double emulsions. PON1 was evolved for the hydrolysis of thiobutyrlactones (TBLs, **13**), a generally poor substrate of PON1 (k_{cat}/K_M = 75 M^{-1} s^{-1}) (Figure 11a). Starting from a library of 10^6 mutants, 3 cycles of screening led to a variant with up to one hundredfold higher TBLase activity ($k_{cat}/K_M = 10^4$ M^{-1} s^{-1}) (Figure 11b).[32] A variant of the same enzyme, rePON1, was further investigated as a target against nerve agents based on organophosphates. Applying random and targeted mutagenesis, coupled to high throughput FACS screening and MTP assays, mutants capable of hydrolyzing cyclosarin with $k_{cat}/K_M \sim 10^7$ M^{-1} s^{-1} were identified. These findings were also applied to prophylactic studies involving mice, where the identified hits exhibited considerable protection against a lethal dose of a cyclosarin derivative.[115] Further applications include the directed evolution of *i) β*-glucosidase leading to a twofold increase in lactose specificity and catalytic turnover rates,[116] *ii)* the development of a model protease with 1.6-fold increased resistance towards the inhibitor antipain

dihydrochloride,[117] and *iii)* the screening of a cellulase mutant library with the identification of variants with over 13-fold increased specific activity compared to wild type.[118]

Prodanovic *et al.* highlighted the versatility of this screening platform by expanding this technology to *in vivo*-encapsulated yeast in combination with a vanadium bromoperoxidase-coupled fluorescence assay (ViPer) to detect H_2O_2 (Scheme 2). In the assay, H_2O_2 was used by the bromoperoxidase to produce hypobromide which reacted with a fluorogenic probe to release fluorescent coumarin. Using this approach, a 200-fold enrichment of active GOx was identified in a single screening round starting from a library of ~10^4 variants.[111] Similarly, cellulase activity was evolved to achieve a 12-fold enrichment of the active variant in a single round.[119]

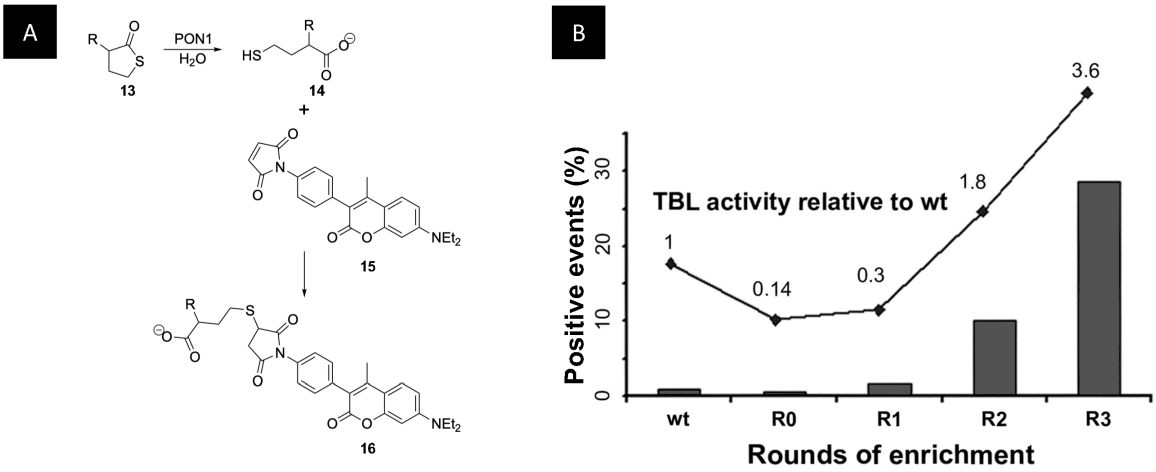
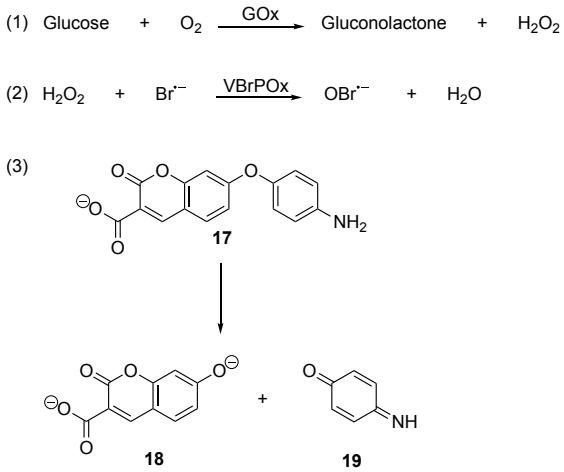


Figure 11. a) PON1 hydrolyzes γTBL (**13**) to the corresponding thiol **14**. N-(4-(7-diethylamino-4-methylcoumarin-3-yl)phenyl)maleimide (CPM, **15**) reacts with the free thiol to form a fluorescent product (**16**). **b)** Development of the fluorescence intensity over three rounds of enrichment. TBLase activity was determined in the crude lysate of the selected pool and normalized to the activity of wild type PON1 (wt).[32] (Figure reprinted with permission from **b)** Elsevier.)



Scheme 2. A vanadium bromoperoxidase-coupled (ViPer) fluorescent assay for the directed evolution of GOx and subsequent sorting.[111]

1.4.2 TECHNOLOGY ADVANCES II: ON-CHIP FORMATION AND STABILITY OPTIMIZATION FOR HIGH-THROUGHPUT SORTING

Producing double emulsions on-chip greatly improved the monodispersity and the control over the number of inner aqueous phase compartments.[120] Furthermore, it greatly improved the droplet sorting efficiency and throughput. However, the production of double emulsion droplets on-chip is more challenging than the formation of single emulsions as it requires different surface wetting properties for each emulsification step. The microfluidic junctions need to be hydrophobic for the first w/o emulsion and hydrophilic for the second w/o/w emulsion. Various strategies have been investigated to address this challenge, such as decoupling the two emulsification steps, using coating solutions or building the chip from different materials. In a pioneer study, Okushima and coworkers proposed several design options allowing for either *i)* decoupled emulsification steps in two quartz and Pyrex glass chips or *ii)* double emulsification on one single Pyrex glass chip. For both designs, the double emulsions were produced using T-junctions. The requirements for different channel surface properties at the junctions were satisfied by coating the first junction hydrophobically with a silane-coupling agent. Using these devices, the authors produced monodisperse double emulsions of about 100 μm in diameter at a rate of 22 Hz (Figure 12b).[121,122] Another study reported the formation of double emulsions at higher throughput using glass microcapillaries. Droplets of 10-50 μm in diameter were produced at rates ranging from 100-5000 Hz. The authors further highlighted the potential of their device by controlling the size of the inner water droplet and oil shell of the double emulsion.[123]

The need for rapid prototyping and simple microfabrication led the field towards the use of PDMS-based microfluidic devices. In a first study on PDMS surface modification using plasma polymerization, the authors achieved selective hydrophilic coating and subsequent formation of double emulsions with a T-junction.[39] The formation of double emulsions with controlled oil shell thickness was reported by Abate *et al.*[40] A PDMS chip with two consecutive flow-focusing junctions was selectively coated using a flow-confinement technique.[124] Similar devices with a step structure at the second flow-focusing junction were developed to facilitate the second emulsification.[125,126] Due to the critical nature of the coating and the precision required for the wettability patterning, different strategies for coating or decoupling the two emulsification steps were developed and are described in detail elsewhere.[127,128]

Double emulsions formed on-chip initially found applications in cell culturing or *in vitro* protein expressions (Figure 12c).[47,129,130] Notably, Zhang and coworkers encapsulated *E. coli* in monodisperse double emulsions produced on two decoupled PDMS chips. They studied bacterial growth and protein expression by addition of the inducer in the outer aqueous phase and utilizing diffusion across the oil layer.[131]

The ability to use FACS on double emulsions constituted an essential advance for improved compatibility of double emulsions in the context of directed evolution studies (Figure 12d). Efforts were therefore invested in studying the deformation of double emulsions in FACS devices and in identifying suitable surfactants to ensure droplet stability.[132,133]

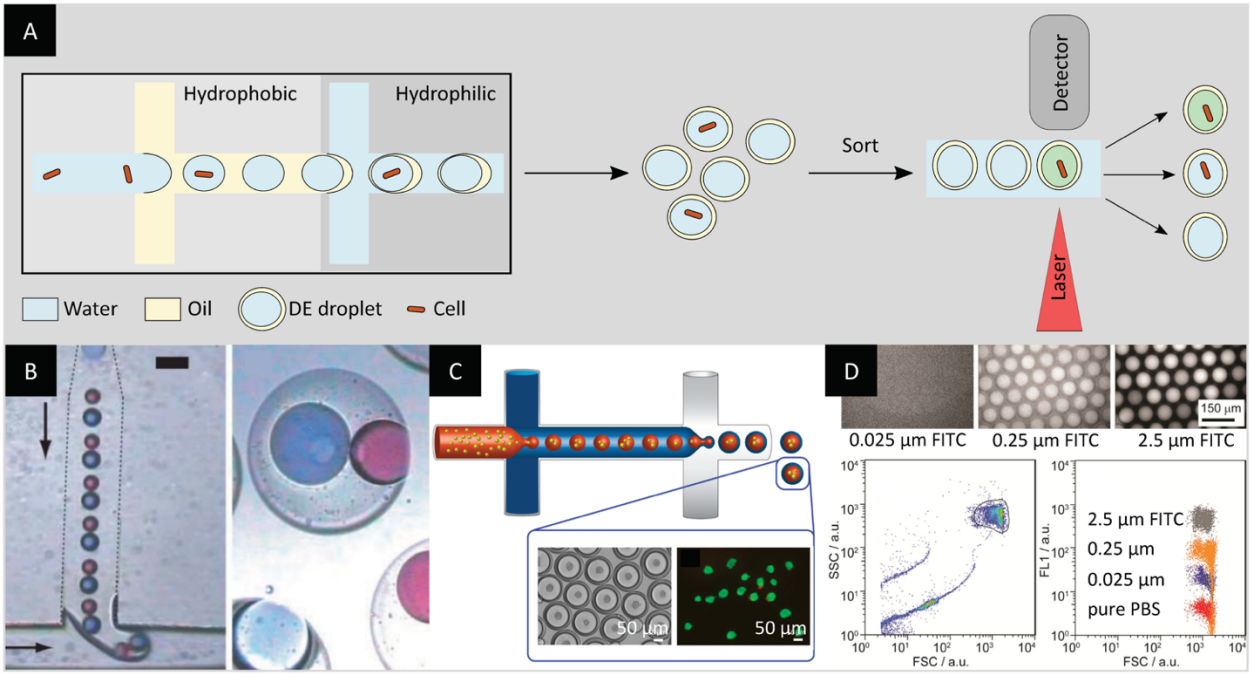


Figure 12. Microfluidics-based formation of double emulsion droplets. a) Double emulsion (DE) droplets are formed in a microfluidic chip using two consecutive flow-focusing junctions. The first emulsification step requires hydrophobic channel walls while the second step requires hydrophilic channels. The DE droplets can be sorted by conventional FACS. **b)** Micrograph of the formation of DE droplets with two aqueous compartments. Scale bar: 100 µm.[122] **c)** DE formation and encapsulation of cells for the growth of multicellular spheroids.[129] **d)** Micrographs of DEs containing discrete concentrations of fluorescent dye and FACS plot displaying the discrimination between the different DE populations.[132] (Images reprinted with permission from **b)** the Royal Society of Chemistry, **c)** Springer Nature Limited. Image adapted **d)** from reference [132].)

1.4.2.1 APPLICATIONS: *IN VITRO* AND *IN VIVO*

These technological advances led to the development of a platform for single cell and enzymatic activity-screening. Terekhov and Smirnov *et al.* combined FACS sorting of double emulsions with downstream next-generation sequencing and liquid chromatography-mass spectrometry (LC-MS) analysis of secretome and proteome. In their comprehensive study, the authors used a two-step on-chip emulsification process to perform enzyme screenings with different organisms. They succeeded in sorting active yeast cells displaying an enzyme on their membrane from a non-active population using a fluorogenic substrate. Several mixing ratios – up to $1{:}10^5$ – were investigated, and the authors achieved maximum enrichments for the low dilutions and significant enrichment for the highest dilution. They further illustrated the potential of their platform for distinguishing between different enzymatic activities and between different levels of enzymatic activity. Finally, the cell-to-cell interaction between different organisms was investigated using yeast and bacterial cells.[134]

First studies highlighting the power of double emulsions include the efficient enrichment of active wild-type arylsulfatase from a low activity mutant. Enrichment factors of 800- and 2500-fold, starting from populations of 0.1% and 0.01% active cells, respectively, have been reported.[135] Using a fluorescent reporter system which gave a positive signal upon full-length amplification of the template DNA by the target polymerase, Larsen *et al.* expanded polymerase function to non-natural genetic polymers. After establishing the approach by enriching a model engineered polymerase ~1200-fold, the screening method was applied to evolve a manganese-independent α-L-threofuranosyl nucleic acid (TNA) polymerase. In merely one round of selection, they identified a manganese independent TNA polymerase with higher fidelity and ~14-fold improved activity.[41]

1.5 LATEST DEVELOPMENTS AND LABEL-FREE METHODS

Recent progress on the microfluidic/technology side focus on optimizing existing tools and methods aiming at a more straightforward use and more reproducible results. Notably, Sukovitch, Kim and co-authors proposed a method to simplify double emulsion production while conserving the monodispersity. They coupled single emulsion production on-chip with a second bulk emulsification step, circumventing the complex coating process required for double emulsion chips.[136] New ways of delivering reagents inside single or double emulsions, mainly by adapting the surfactant type and concentration, have been characterized by several research groups.[137–141] In parallel, substantial advances have been achieved in expanding the screening capabilities of microfluidic platforms. In a groundbreaking study, Ma and coworkers reported a dual-channel microfluidic droplet screening system (DMDS). This system enables the simultaneous sorting of w/o droplets according to two properties of a target enzyme using two different fluorogenic substrates (Figure 13a). The efficiency of the platform for the screening of complex enzymatic properties was illustrated with the directed evolution of a highly enantioselective esterase from *Archeoglobus fulgidus* (AFEST) (Figure 13b).

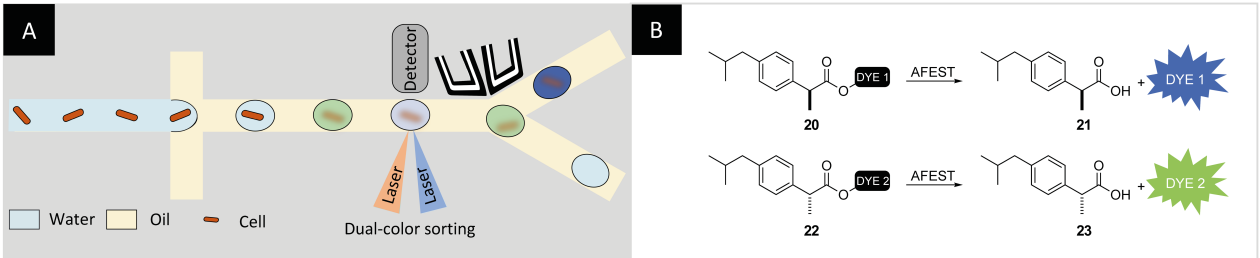


Figure 13. a) Schematic representation of the dual-channel microfluidic droplet screening (DMDS). **b)** (S)-ibuprofen and (R)-ibuprofen modified with two different fluorophores and the enzymatic reaction yielding two different fluorescent signals. Substrate **20** is used as the selection substrate and substrate **22** as the counter-selection substrate.

To improve the enantioselectivity of AFEST towards (S)-ibuprofen, variants with increased fluorescence signal for dye 1 but lower fluorescence signal for dye 2 were selected using on-chip sorting.[142] (Image adapted from reference [142].)

After five rounds of evolution, a variant with 700-fold improvement in enantioselectivity for (*S*)-profens was obtained.[142] In a recent study, Brower *et al.* introduced a comprehensive FACS-based method to sort and isolate double emulsion droplets produced on-chip. Their method allows for the encapsulation of a variety of mammalian cells and sorting at throughputs >10 kHz while maintaining the w/o/w droplets' integrity, followed by retrieval of genetic material.[143,144]

Despite fluorescence detection still being the gold standard for assaying enzymatic activities in droplets, reports investigating other techniques have recently gained attention. These techniques give access to enzyme characteristics without requiring the use of a fluorogenic substrate. A first example reported by Gielen *et al.* introduced a microfluidic device for absorbance-activated droplet sorting (AADS). With this device, the authors evolved a phenylalanine dehydrogenase over two rounds of screening and found a variant with 4.5-fold increased activity and >10ºC increased thermostability.[145] Similarly, passive sorting strategies, such as sorting by interfacial tension, will allow novel types of assays, where changes in droplet content translate into different droplet properties.[94]

One of the most promising and widely applicable alternatives to fluorescence-based readouts is mass spectrometry (MS), which allows label-free multiplexed characterization of several analytes. In the past years, several groups have illustrated the compatibility of droplet microfluidics with Electrospray Ionization (ESI)-MS[146,147] or Matrix-Assisted Laser Desorption/Ionization (MALDI)-MS for the analysis of enzyme secretion of yeast cells.[148] Notably, the Kennedy group has reported the coupling of w/o droplets with high throughput MS for the *in vitro* screening of enzyme inhibitors and activators.[149,150] With their methods, droplets can be directly injected in the ESI-MS at a throughput of almost 1 Hz. In a follow-up study, the authors increased their platform throughput more than 3-fold. The method was applied to the screening of transaminase libraries and further highlighted the compatibility of their system with *in vitro* translation-transcription of proteins (Figure 14a).

The most recent advance in this field concerns the screening of enzymatic reactions with an innovative method, termed MADS (Mass Activated Droplet Sorting). MADS combines MS analysis with FADS and benefits both from the high sensitivity of MS and from the possibility of collecting the sample allowed by FADS. The MADS device allows for droplet production and splitting of each droplet in two. One fraction is analyzed by ESI-MS while the second fraction follows a delay channel leading to a FADS electrode (Figure 14b). The ESI-MS results allow the active sorting of

the second fraction with a throughput of 0.7 Hz (Figure 14c). The authors applied their methods to the activity-based screening of a model transaminase library expressed *in vitro*.[151]

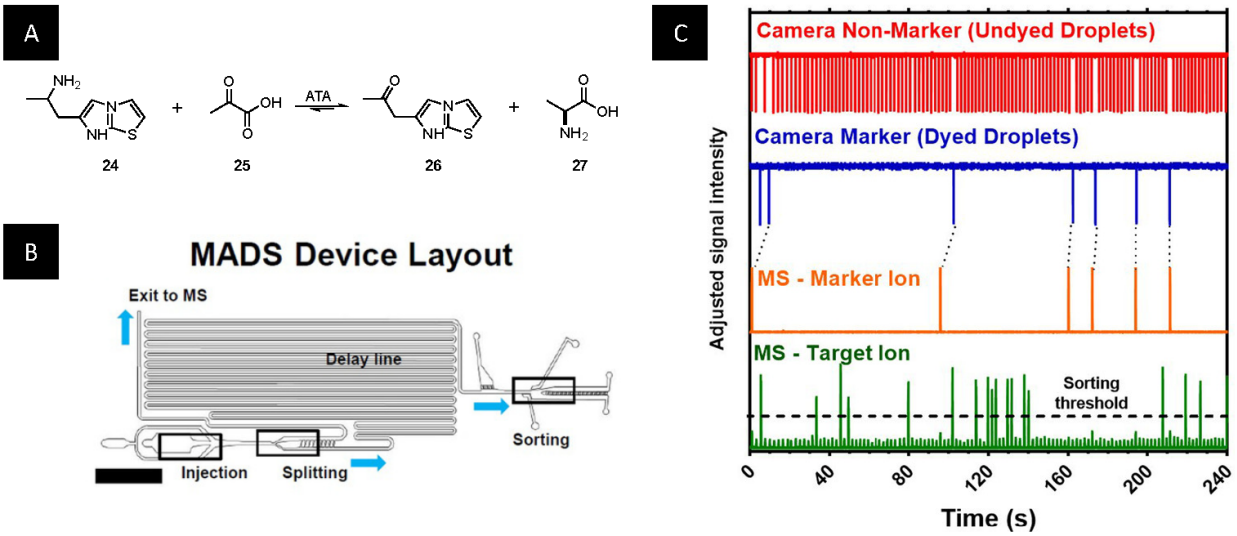


Figure 14. a) The transaminase activity of ATA 117 is screened by evaluating the transformation of the non-native ATA-substrate (**24**) and pyruvate (**25**) to the ATA-product (**26**) and alanine (**27**) after ivTT. **b)** Schematic representation of the Mass Activated Droplet Sorting (MADS) device. Nanoliter-sized droplets are injected in the bottom left region (*'injection'*) and are split asymmetrically (*'splitting'*). While the larger droplet travels directly to the mass spectrometer, the smaller droplet flows through the delay line. The smaller droplet is sorted using a dielectrophoretic sorter (*'sorting'*) according to the sorting decision made using the MS-signal. **a)** For the MADS device to function, three different samples are analyzed in parallel. Inactive and therefore uncolored droplets are recognized by a camera by pattern tracing (red). Marker droplets for synchronization contain food color and are detected by the camera (blue). The signals are synchronized with the MS-signal of the marker-ion (orange). After synchronization, the MS-signal of the target ion (green) is used to make a sorting decision.[151] (Image reprinted with permission from Wiley-WCH Verlag GmbH & Co KGaA.)

1.6 Outlook

Moving away from model sorts and display of platform capabilities, many research groups are now working on the improvement of enzymes with industrial or medical relevance.[65,74,76,150,152,153] Besides the evolution of natural enzymes, the toolkit of directed evolution is expanding to artificial enzymes to introduce non-natural reactivities urgently needed in the pharmaceutical industry,[152] *de novo* designed enzymes to understand and reengineer enzyme active sites and, more recently, machine-learning assisted directed evolution.[154] Furthermore, as fluorogenic substrates cannot always be synthesized, novel strategies for fast, but non-fluorescence-based detection will be critical for future developments.

The advances in droplet microfluidics over the past 20 years have permitted decisive steps towards the discovery of enzymes with new or improved functionalities. The higher throughput and facilitated sorting allowed the directed evolution of libraries of increasing sizes at a significantly reduced time and material consumption. Droplet microfluidics paved the way to automated workflows, faster screenings and enabled the use of conventional FACS, accessible in most biology institutes. Close collaborations between engineering groups and chemistry or biology research

groups showed a synergetic effect by allowing successful large campaigns. To continue on this prosperous avenue, microfluidics systems must be further simplified to enable robust operation of microfluidic devices by non-experts. Cheap, commercially-available microchips will further lower the hurdles to exchange standard tools for microfluidic systems.[155]

With more and more groups working on directed evolution using microfluidic systems, the spectrum of applications and assays will broaden. We believe that the joint effort from these two fields holds great promise, and we are looking forward to see the new innovative developments that will emerge from collaborations between engineers and biochemists in the future!

ACKNOWLEDGEMENTS

The authors thank the Swiss National Science Foundation (Grant 200020_182046), the NCCR Molecular Systems Engineering and the ERC (the DrEAM - Advanced Grant 694424 and the HybCell - Consolidator Grant 681587) for their generous support.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

1.7 REFERENCES

[1] O. Kuchner, F. H. Arnold, *Trends Biotechnol.* **1997**, *15*, 523–530.

[2] A. Li, G. Qu, Z. Sun, M. T. Reetz, *ACS Catal.* **2019**, *9*, 7769–7778.

[3] K. Chen, F. H. Arnold, *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 5618–5622.

[4] G. Qu, A. Li, C. G. Acevedo-Rocha, Z. Sun, M. T. Reetz, *Angew. Chemie - Int. Ed.* **2020**, *59*, 13204–13231.

[5] Z. Zou, D. M. Mate, M. Nöth, F. Jakob, U. Schwaneberg, *Chem. - A Eur. J.* **2020**, *26*, 13568-13574.

[6] M. T. Reetz, A. Zonta, K. Schimossek, K.-E. Jaeger, K. Liebeton, *Angew. Chemie Int. Ed. English* **1997**, *36*, 2830–2832.

[7] I. V. Pavlidis, M. S. Weiß, M. Genz, P. Spurr, S. P. Hanlon, H. Iding, U. T. Bornscheuer, *Nat. Chem.* **2016**, *8*, 1076–1082.

[8] D. S. Macdonald, X. Garrabou, C. Klaus, R. Verez, T. Mori, D. Hilvert, *J. Am. Chem. Soc* **2020**, *142*, 2020.

[9] P. S. Coelho, E. M. Brustad, A. Kannan, F. H. Arnold, *Science.* **2013**, *339*, 307–310.

[10] S. C. Hammer, G. Kubik, E. Watkins, S. Huang, H. Minges, F. H. Arnold, *Science.* **2017**, *358*, 215–218.

[11] O. F. Brandenberg, K. Chen, F. H. Arnold, *J. Am. Chem. Soc.* **2019**, *141*, 8989–8995.

[12] A. Li, C. G. Acevedo-Rocha, L. D'Amore, J. Chen, Y. Peng, M. Garcia-Borràs, C. Gao, J. Zhu, H. Rickerby, S. Osuna, J. Zhou, M. T. Reetz, *Angew. Chemie Int. Ed.* **2020**, DOI 10.1002/anie.202003139.

[13] L. Giger, S. Caner, R. Obexer, P. Kast, D. Baker, N. Ban, D. Hilvert, *Nat. Chem. Biol.* **2013**, *9*, 494–498.

[14] S. Studer, D. A. Hansen, Z. L. Pianowski, P. R. E. Mittl, A. Debon, S. L. Guffy, B. S. Der, B. Kuhlman, D. Hilvert, *Science.* **2018**, *362*, 1285–1288.

[15] A. R. Grimm, D. F. Sauer, M. D. Davari, L. Zhu, M. Bocola, S. Kato, A. Onoda, T. Hayashi, J. Okuda, U. Schwaneberg, *ACS Catal.* **2018**, *8*, 3358–3364.

[16] W. Birmingham, A. Toftgaard Pedersen, M. Dias Gomes, M. Bøje Madsen, M. Breuer, J. Woodley, N. J. Turner, T. Pedersen, D. Gomes, B. Madsen, **2020**, DOI 10.26434/chemrxiv.12925373.v1.

[17] A. P. Green, N. J. Turner, E. O'Reilly, *Angew. Chemie Int. Ed.* **2014**, *53*, 10714–10717.

[18] A. J. Ruff, A. Dennig, G. Wirtz, M. Blanusa, U. Schwaneberg, *ACS Catal.* **2012**, *2*, 2724–2728.

[19] S. Kille, C. G. Acevedo-Rocha, L. P. Parra, Z. G. Zhang, D. J. Opperman, M. T. Reetz, J. P. Acevedo, *ACS Synth. Biol.* **2013**, *2*, 83–92.

[20] A. K. Price, B. M. Paegel, *Anal. Chem.* **2016**, *88*, 339–353.

[21] H. Leemhuis, V. Stein, A. D. Griffiths, F. Hollfelder, *Curr. Opin. Struct. Biol.* **2005**, *15*, 472–478.

[22] A. Rothe, R. N. Surjadi, B. E. Power, *Trends Biotechnol.* **2006**, *24*, 587–592.

[23] U. Markel, K. D. Essani, V. Besirlioglu, J. Schiffels, W. R. Streit, U. Schwaneberg, *Chem. Soc. Rev.* **2020**, *49*, 233–262.

[24] A. Autour, M. Ryckelynck, *Micromachines* **2017**, *8*, 128.

[25] L. A. Kamentsky, *Cytology Automation*, ACADEMIC PRESS, INC., **1971**.

[26] J. Picot, C. L. Guerin, C. Le Van Kim, C. M. Boulanger, *Cytotechnology* **2012**, *64*, 109–130.

[27] S. Fiedler, S. G. Shirley, T. Schnelle, G. Fuhr, *Anal. Chem.* **1998**, *70*, 1909–1915.

[28] M. Eiger, R. Riglert, *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 5740–5747.

[29] N. D. Weiner, *J. Pharm. Sci.* **1975**, *64*, 1434.

[30] F. J. Ghadessy, J. L. Ong, P. Holliger, *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 4552–4557.

[31] M. F. Ficheux, L. Bonakdar, F. Leal-Calderon, J. Bibette, *Langmuir* **1998**, *14*, 2702–2706.

[32] A. Aharoni, G. Amitai, K. Bernath, S. Magdassi, D. S. Tawfik, *Chem. Biol.* **2005**, *12*, 1281–1289.

[33] T. Thorsen, R. W. Roberts, F. H. Arnold, S. R. Quake, *Phys. Rev. Lett.* **2001**, *86*, 4163–4166.

[34] S. L. Anna, N. Bontoux, H. A. Stone, *Appl. Phys. Lett.* **2003**, *82*, 364–366.

[35] A. D. Griffiths, D. S. Tawfik, *EMBO J.* **2003**, *22*, 24–35.

[36] J. C. Baret, O. J. Miller, V. Taly, M. Ryckelynck, A. El-Harrak, L. Frenz, C. Rick, M. L. Samuels, J. B. Hutchison, J. J. Agresti, D. R. Link, D. A. Weitz, A. D. Griffiths, *Lab Chip* **2009**, *9*, 1850–1858.

[37] D. R. Link, E. Grasland-Mongrain, A. Duri, F. Sarrazin, Z. Cheng, G. Cristobal, M. Marquez, D. A. Weitz, *Angew. Chemie - Int. Ed.* **2006**, *45*, 2556–2560.

[38] R. Obexer, M. Pott, C. Zeymer, A. D. Griffiths, D. Hilvert, *Protein Eng. Des. Sel.* **2016**, *29*, 355–366.

[39] V. Barbier, M. Tatoulian, H. Li, F. Arefi-Khonsari, A. Ajdari, P. Tabeling, *Langmuir* **2006**, *22*, 5230–5232.

[40] A. R. Abate, J. Thiele, D. A. Weitz, *Lab Chip* **2011**, *11*, 253–258.

[41] A. C. Larsen, M. R. Dunn, A. Hatch, S. P. Sau, C. Youngbull, J. C. Chaput, *Nat. Commun.* **2016**, *7*, 1–9.

[42] L. Weng, J. E. Spoonamore, *Micromachines* **2019**, *10*, 734.

[43] P. Becher, *Emulsions: Theory and Practice.*, Reinhold Publishing Corporation, N.Y., **1957**.

[44] I. Burgaud, E. Dickinson, P. V. Nelson, *Int. J. Food Sci. Technol.* **1990**, *25*, 39–46.

[45] W. D. Pandolfe, *J. Dispers. Sci. Technol.* **1995**, *16*, 633–650.

[46] J. Floury, A. Desrumaux, J. Lardières, *Innov. Food Sci. Emerg. Technol.* **2000**, *1*, 127–134.

[47] H. N. Joensson, H. Andersson Svahn, *Angew. Chemie - Int. Ed.* **2012**, *51*, 12176–12192.

[48] A. Sepp, D. S. Tawfik, A. D. Griffiths, *FEBS Lett.* **2002**, *532*, 455–458.

[49] R. Nir, R. Lamed, L. Gueta, E. Sahar, *Appl. Environ. Microbiol.* **1990**, *56*, 2870–2875.

[50] J. C. Weaver, J. G. Bliss, K. T. Powell, G. I. Harrison, G. B. Williams, **1991**, *9*, 873–877.

[51] E. Sahar, R. Nir, R. Lamed, *Cytometry* **1994**, *15*, 213–221.

[52] A. D. Griffiths, D. S. Tawfik, *Nat. Biotechnol.* **1998**, *16*, 652–656.

[53] Y.-F. Lee, D.-S. Tawfik, A.-D. Griffiths, *Nucleic Acids Res.* **2002**, *30*, 4937–4944.

[54] J. J. Agresti, B. T. Kelly, A. Jäschke, A. D. Griffiths, *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 16170–16175.

[55] B. M. Paegel, G. F. Joyce, *Chem. Biol.* **2010**, *17*, 717–724.

[56] J. J. Agresti, B. T. Kelly, A. Jäschke, A. D. Griffiths, *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 16170–16175.

[57] B. Arezi, N. McKinney, C. Hansen, M. Cayouette, J. Fox, K. Chen, J. Lapira, S. Hamilton, H. Hogrefe, *Front. Microbiol.* **2014**, *5*, 1–10.

[58] R. Prodanovic, R. Ostafe, A. Scacioc, U. Schwaneberg, *Comb. Chem. High Throughput Screen.* **2012**, *14*, 55–60.

[59] R. Ostafe, R. Prodanovic, J. Nazor, R. Fischer, *Chem. Biol.* **2014**, *21*, 414–421.

[60] G. Kovačević, R. Ostafe, A. M. Balaž, R. Fischer, R. Prodanović, *J. Biosci. Bioeng.* **2019**, *127*, 30–37.

[61] J. A. Stapleton, J. R. Swartz, *PLoS One* **2010**, *5*, 1–8.

[62] B. Zhu, T. Mizoguchi, T. Kojima, H. Nakano, *PLoS One* **2015**, *10*, 1–17.

[63] M. Levy, K. E. Griswold, A. D. Ellington, *RNA* **2005**, *11*, 1555–1562.

[64] P. Gianella, E. L. Snapp, M. Levy, *Biotechnol. Bioeng.* **2016**, *113*, 1647–1657.

[65] C. Femmer, M. Bechtold, M. Held, S. Panke, *Metab. Eng.* **2020**, *59*, 15–23.

[66] S. Y. Teh, R. Lin, L. H. Hung, A. P. Lee, *Lab Chip* **2008**, *8*, 198–220.

[67] G. M. Whitesides, *Nature* **2006**, *442*, 368–373.

[68] H. Song, D. L. Chen, R. F. Ismagilov, *Angew. Chemie - Int. Ed.* **2006**, *45*, 7336–7356.

[69] J. M. Köhler, T. Henkel, A. Grodrian, T. Kirner, M. Roth, K. Martin, J. Metze, *Chem. Eng. J.* **2004**, *101*, 201–216.

[70] J. F. Edd, D. Di Carlo, K. J. Humphry, S. Köster, D. Irimia, D. A. Weitz, M. Toner, *Lab Chip* **2008**, *8*, 1262–1264.

[71] M. Chabert, J. L. Viovy, *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 3191–3196.

[72] F. Courtois, L. F. Olguin, G. Whyte, A. B. Theberge, W. T. S. Huck, F. Hollfelder, C. Abell, *Anal. Chem.* **2009**, *81*, 3008–3016.

[73] M. Fischlechner, Y. Schaerli, M. F. Mohamed, S. Patil, C. Abell, F. Hollfelder, *Nat. Chem.* **2014**, *6*, 791–796.

[74] T. C. Scanlon, S. M. Dostal, K. E. Griswold, *Biotechnol. Bioeng.* **2014**, *111*, 232–243.

[75] J. M. Duarte, I. Barbier, Y. Schaerli, *ACS Synth. Biol.* **2017**, *6*, 1988–1995.

[76] C. Ma, Z. L. Tan, Y. Lin, S. Han, X. Xing, C. Zhang, *J. Biosci. Bioeng.* **2019**, *128*, 662–668.

[77] N. Bremond, J. Bibette, *Soft Matter* **2012**, *8*, 10549–10559.

[78] M. Chabert, K. D. Dorfman, J. L. Viovy, *Electrophoresis* **2005**, *26*, 3706–3715.

[79] K. Ahn, J. Agresti, H. Chong, M. Marquez, D. A. Weitz, *Appl. Phys. Lett.* **2006**, *88*, 264105.

[80] A. R. Abate, T. Hung, P. Marya, J. J. Agresti, D. A. Weitz, *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 19163–19166.

[81] A. Sciambi, A. R. Abate, *Biomicrofluidics* **2013**, *7*, 44112.

[82] A. M. Nightingale, T. W. Phillips, J. H. Bannock, J. C. De Mello, *Nat. Commun.* **2014**, *5*, 1–8.

[83] T. Taniguchi, T. Torii, T. Higuchi, *Lab Chip* **2002**, *2*, 19–23.

[84] K. Ahn, C. Kerbage, T. P. Hunt, R. M. Westervelt, D. R. Link, D. A. Weitz, *Appl. Phys. Lett.* **2006**, *88*, 1–3.

[85] P. S. Dittrich, A. Manz, *Anal. Bioanal. Chem.* **2005**, *382*, 1771–1782.

[86] P. S. Dittrich, M. Jahnz, P. Schwille, *ChemBioChem* **2005**, *6*, 811–814.

[87] L. Mazutis, J. C. Baret, P. Treacy, Y. Skhiri, A. F. Araghi, M. Ryckelynck, V. Taly, A. D. Griffiths, *Lab Chip* **2009**, *9*, 2902–2908.

[88] S. L. Sjostrom, Y. Bai, M. Huang, Z. Liu, J. Nielsen, H. N. Joensson, H. Andersson Svahn, *Lab Chip* **2014**, *14*, 806–813.

[89] O. Caen, S. Schütz, M. S. S. Jammalamadaka, J. Vrignon, P. Nizard, T. M. Schneider, J. C. Baret, V. Taly, *Microsystems Nanoeng.* **2018**, *4*, 264105.

[90] A. Fallah-Araghi, J. C. Baret, M. Ryckelynck, A. D. Griffiths, *Lab Chip* **2012**, *12*, 882–891.

[91] R. Ostafe, R. Prodanovic, W. L. Ung, D. A. Weitz, R. Fischer, *Biomicrofluidics* **2014**, *8*, 3–6.

[92] D. Frenzel, C. A. Merten, *Lab Chip* **2017**, *17*, 1024–1030.

[93] S. Hasan, D. Geissler, K. Wink, A. Hagen, J. J. Heiland, D. Belder, *Lab Chip* **2019**, *19*, 403–409.

[94] C. W. Pan, D. G. Horvath, S. Braza, T. Moore, A. Lynch, C. Feit, P. Abbyad, *Lab Chip* **2019**, *19*, 1344–1351.

[95] N. Nitta, T. Sugimura, A. Isozaki, H. Mikami, K. Hiraki, S. Sakuma, T. Iino, F. Arai, T. Endo, Y. Fujiwaki, H. Fukuzawa, M. Hase, T. Hayakawa, K. Hiramatsu, Y. Hoshino, M. Inaba, T. Ito, H. Karakawa, Y. Kasai, K. Koizumi, S. W. Lee, C. Lei, M. Li, T. Maeno, S. Matsusaka, D. Murakami, A. Nakagawa, Y. Oguchi, M. Oikawa, T. Ota, K. Shiba, H. Shintaku, Y. Shirasaki, K. Suga, Y. Suzuki, N. Suzuki, Y. Tanaka, H. Tezuka, C. Toyokawa, Y. Yalikun, M. Yamada, M. Yamagishi, T. Yamano, A. Yasumoto, Y. Yatomi, M. Yazawa, D. Di Carlo, Y. Hosokawa, S. Uemura, Y. Ozeki, K. Goda, *Cell* **2018**, *175*, 266-276.e13.

[96] V. Anagnostidis, B. Sherlock, J. Metz, P. Mair, F. Hollfelder, F. Gielen, *Lab Chip* **2020**, *20*, 889–900.

[97] H. Goto, Y. Kanai, A. Yotsui, S. Shimokihara, S. Shitara, R. Oyobiki, K. Fujiwara, T. Watanabe, Y. Einaga, Y. Matsumoto, N. Miki, N. Doi, *Lab Chip* **2020**, *20*, 852–861.

[98] J. J. Agresti, E. Antipov, A. R. Abate, K. Ahn, A. C. Rowat, J. C. Baret, M. Marquez, A. M. Klibanov, A. D. Griffiths, D. A. Weitz, *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 4004–4009.

[99] B. Kintses, C. Hein, M. F. Mohamed, M. Fischlechner, F. Courtois, C. Lainé, F. Hollfelder, *Chem. Biol.* **2012**, *19*, 1001–1009.

[100] P. A. Romero, T. M. Tran, A. R. Abate, *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 7159–7164.

[101] R. Obexer, M. Pott, C. Zeymer, A. D. Griffiths, D. Hilvert, *Protein Eng. Des. Sel.* **2016**, *29*, 355–366.

[102] R. Obexer, A. Godina, X. Garrabou, P. R. E. Mittl, D. Baker, A. D. Griffiths, D. Hilvert, *Nat. Chem.* **2017**, *9*, 50–56.

[103] A. Debon, M. Pott, R. Obexer, A. P. Green, L. Friedrich, A. D. Griffiths, D. Hilvert, *Nat. Catal.* **2019**, *2*, 740–747.

[104] G. Yang, S. G. Withers, *ChemBioChem* **2009**, *10*, 2704–2715.

[105] A. K. Price, B. M. Paegel, *Anal. Chem.* **2016**, *88*, 339–353.

[106] F. W. Y. Chiu, S. Stavrakis, *Electrophoresis* **2019**, *40*, 2860–2872.

[107] N. Garti, C. Bisperink, *Curr. Opin. Colloid Interface Sci.* **1998**, *3*, 657–667.

[108] A. Benichou, A. Aserin, N. Garti, *Polym. Adv. Technol.* **2002**, *13*, 1019–1031.

[109] V. Muguet, M. Seiller, G. Barratt, O. Ozer, J. P. Marty, J. L. Grossiord, *J. Control. Release* **2001**, *70*, 37–49.

[110] C. Goubault, K. Pays, D. Olea, P. Gorria, J. Bibette, V. Schmitt, F. Leal-Calderon, *Langmuir* **2001**, *17*, 5184–5188.

[111] R. Prodanovic, R. Ostafe, M. Blanusa, U. Schwaneberg, *Anal. Bioanal. Chem.* **2012**, *404*, 1439–1447.

[112] F. Ma, Y. Xie, C. Huang, Y. Feng, G. Yang, *PLoS One* **2014**, *9*, e89785.

[113] K. Bernath, M. Hai, E. Mastrobattista, A. D. Griffiths, S. Magdassi, D. S. Tawfik, *Anal. Biochem.* **2004**, *325*, 151–157.

[114] E. Mastrobattista, V. Taly, E. Chanudet, P. Treacy, B. T. Kelly, A. D. Griffiths, *Chem. Biol.* **2005**, *12*, 1291–1300.

[115] R. D. Gupta, M. Goldsmith, Y. Ashani, Y. Simo, G. Mullokandov, H. Bar, M. Ben-David, H. Leader, R. Margalit, I. Silman, J. L. Sussman, D. S. Tawfik, *Nat. Chem. Biol.* **2011**, *7*, 120–125.

[116] E. Hardiman, M. Gibbs, R. Reeves, P. Bergquist, *Appl. Biochem. Biotechnol.* **2010**, *161*, 301–312.

[117] R. Tu, R. Martinez, R. Prodanovic, M. Klein, U. Schwaneberg, *J. Biomol. Screen.* **2011**, *16*, 285–294.

[118] G. Körfer, C. Pitzler, L. Vojcic, R. Martinez, U. Schwaneberg, *Sci. Rep.* **2016**, *6*, 1–12.

[119] R. Ostafe, R. Prodanovic, U. Commandeur, R. Fischer, *Anal. Biochem.* **2013**, *435*, 93–98.

[120] T. Nisisako, *Micro Process Eng. A Compr. Handb.* **2013**, *2*, 345–361.

[121] S. Okushima, T. Nisisako, T. Torii, T. Higuchi, *Langmuir* **2004**, *20*, 9905–9908.

[122] T. Nisisako, S. Okushima, T. Torii, *Soft Matter* **2005**, *1*, 23–27.

[123] A. S. Utada, E. Lorenceau, D. R. Link, P. D. Kaplan, H. A. Stone, D. A. Weitz, *Science.* **2005**, *308*, 537–541.

[124] A. R. Abate, J. Thiele, M. Weinhart, D. A. Weitz, *Lab Chip* **2010**, *10*, 1774–1776.

[125] D. Saeki, S. Sugiura, T. Kanamori, S. Sato, S. Ichikawa, *Lab Chip* **2010**, *10*, 357–362.

[126] F. C. Chang, Y. C. Su, *J. Micromechanics Microengineering* **2008**, *18*, 065018–0655026.

[127] K. Doufène, C. Tourné-Péteilh, P. Etienne, A. Aubert-Pouëssel, *Langmuir* **2019**, *35*, 12597–12612.

[128] B. Thompson, N. Movsesian, C. Cheng, P. Karandikar, M. Gupta, N. Malmstadt, *Modular Microfluidics for Double Emulsion Formation*, Elsevier Inc., **2018**.

[129] H. F. Chan, Y. Zhang, Y. P. Ho, Y. L. Chiu, Y. Jung, K. W. Leong, *Sci. Rep.* **2013**, *3*, 1–8.

[130] N. Wu, J. G. Oakeshott, C. J. Easton, T. S. Peat, R. Surjadi, Y. Zhu, *J. Micromechanics Microengineering* **2011**, *21*, 054032.

[131] Y. Zhang, Y. P. Ho, Y. L. Chiu, H. F. Chan, B. Chlebina, T. Schuhmann, L. You, K. W. Leong, *Biomaterials* **2013**, *34*, 4564–4572.

[132] J. Yan, W. A. C. Bauer, M. Fischlechner, F. Hollfelder, C. F. Kaminski, W. T. S. Huck, *Micromachines* **2013**, *4*, 402–413.

[133] S. Ma, W. T. S. Huck, S. Balabani, *Lab Chip* **2015**, *15*, 4291–4301.

[134] S. S. Terekhov, I. V. Smirnov, A. V. Stepanova, T. V. Bobik, Y. A. Mokrushina, N. A. Ponomarenko, A. A. Belogurov, M. P. Rubtsova, O. V. Kartseva, M. O. Gomzikova, A. A. Moskovtsev, A. S. Bukatin, M. V. Dubina, E. S. Kostryukova, V. V. Babenko, M. T. Vakhitova, A. I. Manolov, M. V. Malakhova, M. A. Kornienko, A. V. Tyakht, A. A. Vanyushkina, E. N. Ilina, P. Masson, A. G. Gabibov, S. Altman, *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 2550–2555.

[135] A. Zinchenko, S. R. A. Devenish, B. Kintses, P. Y. Colin, M. Fischlechner, F. Hollfelder, *Anal. Chem.* **2014**, *86*, 2526–2533.

[136] D. J. Sukovich, S. C. Kim, N. Ahmed, A. R. Abate, *Analyst* **2017**, *142*, 4618–4622.

[137] P. Gruner, B. Riechers, B. Semin, J. Lim, A. Johnston, K. Short, J. C. Baret, *Nat. Commun.* **2016**, *7*, 10392.

[138] G. Etienne, A. Vian, M. Biočanin, B. Deplancke, E. Amstad, *Lab Chip* **2018**, *18*, 3903–3912.

[139] G. Etienne, M. Kessler, E. Amstad, *Macromol. Chem. Phys.* **2017**, *218*, 1–10.

[140] B. Cai, T. Ji, N. Wang, X. Li, R. He, W. Liu, G. Wang, X. Zhao, L. Wang, Z. Wang, *Lab a Chip* -

Miniaturisation Chem. Biol. **2019**, *19*, 422–431.

[141] F. Ma, M. Fischer, Y. Han, S. G. Withers, Y. Feng, G. Y. Yang, *Anal. Chem.* **2016**, *88*, 8587–8595.

[142] F. Ma, M. T. Chung, Y. Yao, R. Nidetz, L. M. Lee, A. P. Liu, Y. Feng, K. Kurabayashi, G. Y. Yang, *Nat. Commun.* **2018**, *9*, 1030.

[143] K. K. Brower, C. Carswell-Crumpton, S. Klemm, B. Cruz, G. Kim, S. G. K. Calhoun, P. M. Fordyce, *Lab Chip* **2020**, *20*, 2062–2074.

[144] K. K. Brower, M. Khariton, P. H. Suzuki, C. Still, G. Kim, S. G. K. Calhoun, L. S. Qi, B. Wang, P. M. Fordyce, *Anal. Chem.* **2020**, *92*, 13262–13270.

[145] F. Gielen, R. Hours, S. Emond, M. Fischlechner, U. Schell, F. Hollfelder, *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, E7383–E7389.

[146] R. T. Kelly, J. S. Page, I. Marginean, K. Tang, R. D. Smith, *Angew. Chemie - Int. Ed.* **2009**, *48*, 6832–6835.

[147] L. M. Fidalgo, G. Whyte, B. T. Ruotolo, J. L. P. Benesch, F. Stengel, C. Abell, C. V. Robinson, W. T. S. Huck, *Angew. Chemie - Int. Ed.* **2009**, *48*, 3665–3668.

[148] D. Haidas, S. Bachler, M. Köhler, L. M. Blank, R. Zenobi, P. S. Dittrich, *Anal. Chem.* **2019**, *91*, 2066–2073.

[149] S. Sun, R. T. Kennedy, *Anal. Chem.* **2014**, *86*, 9309–9314.

[150] S. Sun, B. C. Buer, E. N. G. Marsh, R. T. Kennedy, *Anal. Methods* **2016**, *8*, 3458–3465.

[151] D. A. Holland-Moritz, M. K. Wismer, B. F. Mann, I. Farasat, P. Devine, E. D. Guetschow, I. Mangion, C. J. Welch, J. C. Moore, S. Sun, R. T. Kennedy, *Angew. Chemie - Int. Ed.* **2020**, *59*, 4470–4477.

[152] V. Sabatino, J. G. Rebelein, T. R. Ward, *J. Am. Chem. Soc* **2019**, *141*, 24-30.

[153] J. G. Rebelein, Y. Cotelle, B. Garabedian, T. R. Ward, *ACS Catal.* **2019**, *9*, 4173–4178.

[154] Z. Wu, S. B. Jennifer Kan, R. D. Lewis, B. J. Wittmann, F. H. Arnold, *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 8852–8858.

[155] Y. Ding, P. D. Howes, A. J. Demello, *Anal. Chem.* **2019**, *92*, 132–149.

2

ARTIFICIAL METALLOENZYMES BASED ON THE BIOTIN-STREPTAVIDIN TECHNOLOGY

Artificial metalloenzymes (ArMs) are an exciting and challenging field for the high-throughput *in vivo* directed evolution. Progresses in protein engineering, DNA sequencing and bioinformatics, have had a tremendous impact on the field.[156] Nowadays, any enzyme can be engineered for any desired process. Especially directed evolution and the associated endless possibilities of protein engineering opened up possibilities that were never possible before. In this way, completely new biosynthetic pathways could be created and before unattainable substrates are now accessible.[157] Moreover, enzymes are nowadays being optimized to a point where they can accept unnatural substrates and even work in organic solvents.[158]

These reactions, often also bioorthogonal reactions, show high potential for various applications. A bioorthogonal reaction can be described as a reaction that can be carried out in complex biological systems such as cell cultures, live *E. coli* or even mammalian cells. This requires high biocompatibility which is challenging: the catalysts need to be oxygen and water stable to function under physiological conditions, and must not react with existing biomolecules in the cellular matrix to avoid mutual deactivation.[159] Moreover, it should be a very selective reaction that does not interfere with the cellular matrix. This is a big challenge and although numerous reactions would be interesting and would provide means to synthesize specialty chemicals in the pharmaceutical and fragrance industry, chemical biology, or even human medical therapy, only a handful reactions are known to be compatible under physiological conditions.

The main reasons for the limited number of applications may be traced back to the following: (i) many catalytically-interesting reactions are toxic for living organisms as they are based on cytotoxic metals or require harsh reaction conditions, involving high temperatures and/or organic solvents; (ii) the few known catalysts are inefficient in the cellular environment (slow kinetics, few catalytic turnovers and limited catalyst lifetime); and (iii) the reaction conditions within the biological systems (presence of water, oxygen and thiols) are incompatible with the aforementioned metal-based catalysts.

Artificial metalloenzymes offer a potential solution to this challenge, through the merger of two complementary fields: homogeneous and enzymatic catalysis. The introduction of an abiotic artificial cofactor within a protein may catalyze new-to-nature and bioorthogonal reactions.[160] Homogeneous catalysis often relies on metal catalysts which have a broad reaction scope, but are limited in terms of selectivity, meanwhile enzymatic catalysis shows high efficiencies and exquisite selectivities under comparatively mild reaction conditions (water, pH=7.4, ambient temperature).[161] This high efficiency and selectivity is mainly the result of natural selection over a long period of time. Thousands and thousands of generations allowed the optimization of an active site for a very specific reaction. A prominent example is the orotidine 5'-phophate (OMP) decarboxylase, where the non-catalyzed reaction is estimated to have a half-time ($t_{1/2}$) of 78 million years and an enzyme catalyzed improved the reaction rate of a factor of 10^{17}.[162] A key advantage of enzymes is the stabilization of the transition state by an array of different interactions: hydrogen bonding, electrostatic forces and van der Waals interactions between the substrate, the cofactor and the side chains of the protein allow the tight modulation of the active site, increasing the effective concentration.[163–165] These features, combined with the potential to introduce non-natural reactivities, render ArMs an attractive alternative to traditional approaches.[166] Moreover, ArMs offer two handles to optimize their reactivity by either chemical optimization of the metal cofactor or by genetic engineering of the protein scaffold.[167–169]

The following chapter aims to provide a general concept of ArMs (Chapter 2.1) and an overview of directed evolution campaigns using the biotin-streptavidin technology, while giving an overview of different Sav scaffolds (Chapter 2.3). Finally, the catalytic system based on a pianostool ruthenium complex used in this thesis is introduced (Chapter 0).

Parts of the following section have been published in
CHIMIA, ***2021***, *75*, 257-260.
doi: 10.2533/chimia.2021.257

Towards the Directed Evolution of Artificial Metalloenzymes

Jaicy Vallapurackal*

SCS-Metrohm Award for best oral presentation in Catalysis Science & Engineering (Runner-up)

Department of Chemistry, University of Basel, Mattenstrasse 24a, CH-4058 Basel, Switzerland.
E-mail: jaicy.vallapurackal@unibas.ch

KEYWORDS: Directed evolution · Artificial metalloenzymes · Screening · Droplet Microfluidics

2.1 Introduction

Incorporation of a synthetic metal cofactor within a protein scaffold enables the formation of artificial metalloenzymes (ArMs). In this strategy, the metal catalyst provides the first-coordination sphere with non-natural reactivities and larger substrate scopes. The host protein offers the second coordination sphere, increasing activity and selectivity through the protein environment.

ArMs are formed by mainly three different anchoring strategies: covalent, dative, and supramolecular anchoring. In the covalent approach, a functional group on the cofactor makes an irreversible bond with an amino acid side chain within the protein (Figure 15a). For the dative anchoring, electrostatic interactions between the metal cofactor and the protein are exploited (Figure 15b). Supramolecular anchoring, on the other hand, employs the high affinity of certain molecules for their target enzymes (e.g., biotin-streptavidin). Hereby, the metal cofactor is covalently linked to the high-affinity molecule before it is incorporated into the target enzyme (Figure 15c).[168]

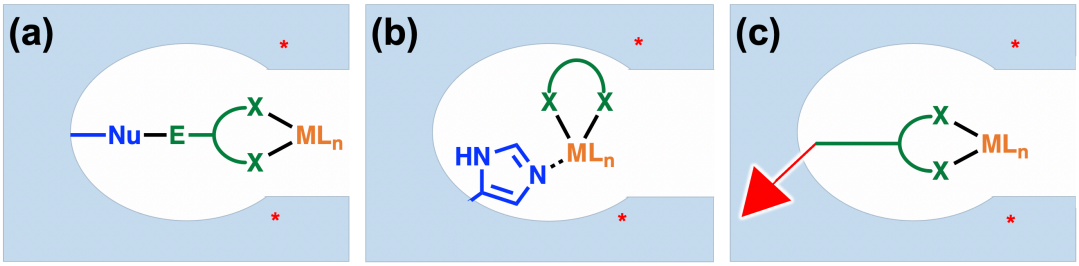


Figure 15. Anchoring strategies to form ArMs. a) Covalent anchoring between a nucleophilic amino acid side chain and an electrophilic functional group of the metal-cofactor. **b)** Dative anchoring using electrostatic interactions between an amino acid and the metal. **c)** Supramolecular anchoring exploiting high affinity of certain molecules for their target protein. (*) genetic modification.

Such combination of a traditional homogeneous catalyst with a protein offers numerous advantages. Catalysis formerly limited to organic solvents become available in physiological conditions, facilitating the implementation of non-natural reactivities *in vivo*. The catalyst is often shielded from the solvent by the deep incorporation of the metallocofactor inside the protein, extending its life-time. A well-defined second coordination sphere, often well-characterized by X-ray crystallography, is an additional advantage. Above all, the protein environment offers a handle for genetic optimization (Figure 15), ultimately enabling the Darwinian-type evolution of a completely synthetic construct.

2.2 The Biotin–Streptavidin Technology

The basic concept of ArMs was introduced almost half a decade ago by Kaiser and Yamamura in 1976 on the example of the Cu^{II} carboxypeptidase A where they exchanged the native zinc ion for a copper ion to achieve oxidase activity.[170] The first incorporation of an abiotic cofactor within a protein scaffold was achieved by Wilson and Whitesides in 1978. They employed the high affinity

of biotin for avidin to anchor Rh(I)-diphosphine and perform hydrogenation of an α-acetamidoacrylic acid in an aqueous environment. This was not only the first proof-of-concept for supramolecular anchoring to build an artificial metalloenzyme, but also the first example for asymmetric catalysis using an ArM. The developed ArM was significantly more efficient and showed moderate stereoselectivity (40% *ee*) when compared to the free metallocofactor in aqueous solution.[171]

Inspired by these results, the Ward laboratory developed ArMs based on the biotin-streptavidin technology, relying on supramolecular anchoring of biotin within streptavidin (Sav). Sav is a homotetrameric protein from the bacterium *Streptomyces avidinii*, showing high affinity towards biotin ($K_d \sim 10^{-14}$ M) and being able to bind up to four molecules of biotin per tetramer. The high affinity is the result of a network of hydrogen bonds provided by the protein side chains that interact with the carboxyl group of biotin (N23, S27, Y43), with its nitrogen (S45, D128) and sulfur (T90) atoms, and with the valeric acid moiety (N49, S88) (Figure 16d). Three tryptophan residues, W79, W92, W102 and W120 from the adjacent monomer, have hydrophobic reactions that further anchor the biotin.[172,173]

It is a highly stable protein with melting temperatures of 73 °C for the apo enzyme and 112 °C for the holo enzyme. It can be lyophilized and tolerates denaturing agents such as guanidinium chloride (GuCl) and urea at a wide range of pH, making it an easy-to-handle protein.[174] Furthermore, it was shown that Sav can tolerate extensive mutations, starting from point mutations to insertions of whole new loop regions. The active site of Sav consists of four loops between the β–sheets $\beta_{3,4}$, $\beta_{4,5}$, $\beta_{5,6}$ and $\beta_{7,8}$. This provides a defined three-dimensional cavity which can accommodate a biotinylated complex and its substrate (Figure 16). As the pocket is formed by loops, the cavity is rather dynamic, which allows fluctuations and gives enough room for the catalysis to take place.[175,176] To date, over a dozen reactions based on this technology have been reported, among which: metathesis,[152] transfer hydrogenation,[177] Suzuki cross-coupling,[178] allylic substitution[179] and more recently hydroxylation.[180]

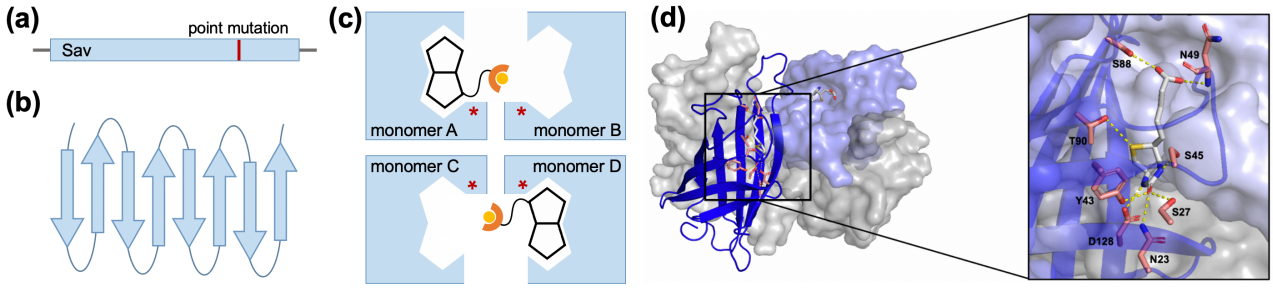


Figure 16. Monomeric Sav. a) Schematic representation of the Sav coding region. **b)** Schematic representation of the Sav secondary structure consisting of eight β-sheets. **c)** Schematic representation of a Sav-tetramer with two incorporated biotinylated metallocofactors. Point mutations are marked in red. **d)** Crystal structure of Sav with biotin incorporated (PDB: 1STP). Three units are displayed as surface in blue, white and grey. One monomer is represented as a cartoon in blue. Positions interacting with biotin (N23, S27, Y43, S45, N49, S88, T90, D128, red) and the hydrogen

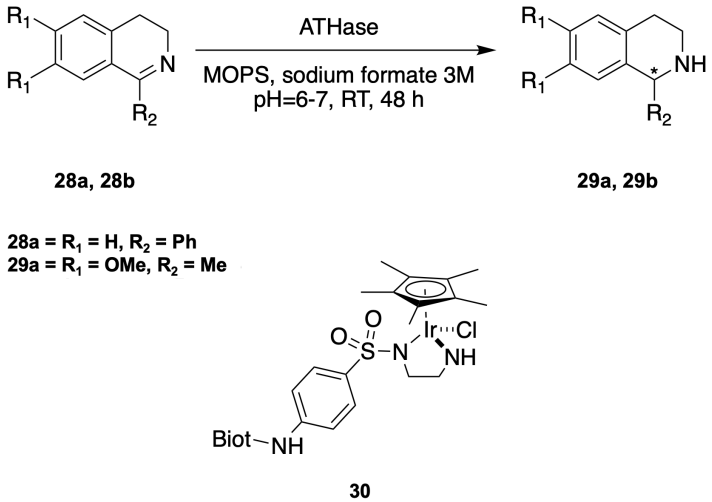
bonding interactions are highlighted (yellow). Only biotin incorporated in the blue monomers are shown and are represented as sticks. Color code: C = grey, N = blue, O = red, S = yellow.

2.3 DIRECTED EVOLUTION OF ARMS

Optimization of ArMs offers mainly two different approaches: either chemical modification to optimize the catalyst or genetic evolution to change the protein's properties. Directed evolution of enzymes is a method providing control to alter biological systems by genetic means allowing to finetune biological activities to optimize certain phenotypes.[181] Frances Arnold was awarded the Nobel prize in 2018 for the directed evolution of enzymes, jointly with George P. Smith and Sir Gregory P. Winter who were honored for their work on phage display of peptides and antibodies. Selected examples of directed evolution performed on ArMs using the biotin-streptavidin technology are described in detail below.

2.3.1 CYTOSOLIC EXPRESSION OF SAV

Different strategies were followed to develop a robust screening platform for the directed evolution of ArMs based on the biotin-streptavidin technology. Initial studies were based on an artificial transfer hydrogenase (ATHase) for the asymmetric hydrogenation of acetamidoacrylate. A small set of purified Sav isomers were screened, and improvements in conversion (quantitative conversion, 100 TONs) and differences between mutants in enantioselectivity (up to 96% *ee*) were observed.[182] However, obtaining lyophilized purified Sav is a time-consuming and laborious process and the screening is strongly limited by the number of variants that can be purified.



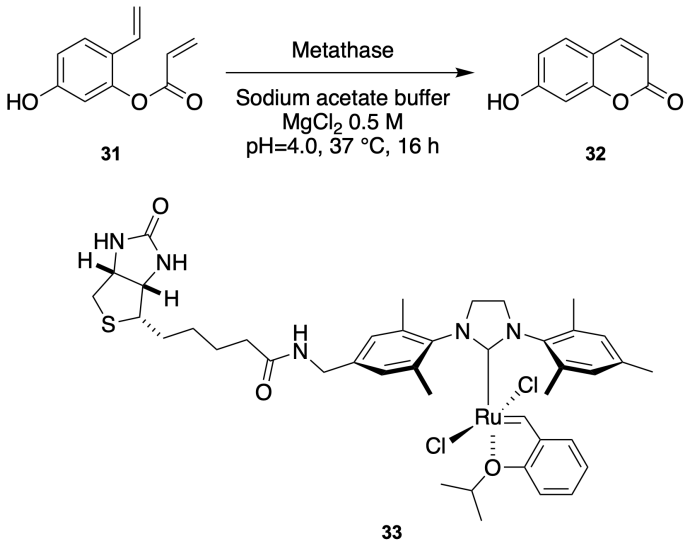
Scheme 3. Reduction of cyclic imines by an ATHase using the biotinylated Ir-based cofactor [Cp*Ir(Biot-*p*-L)Cl].

Applying *in vivo* directed evolution would highly increase the throughput of the screening but the incompatibility of the metal catalyst with the cellular environment makes it challenging. This

incompatibility is mainly due to glutathione (GSH), a thiol present inside cells, which can poison the metal-catalyst. It was observed that the addition of diamide (1,1 azobis(*N*, *N*-dimethyl-formamide) to cell free extracts (CFEs) prevents the poisoning of the Ir-based catalysts for transfer hydrogenation.[183] Based on these results, a directed evolution campaign of an artificial transfer hydrogenase was carried out. Four positions in close vicinity to the embedded metallocofactor [Cp*Ir(Biot-*p*-L)Cl] were selected and screened for the transfer hydrogenation of cyclic imines (Scheme 3). After iterative screening, the two best mutants led to high enantioselectivities of the (*R*)-enantiomer (95% *ee*) and (*S*)-enantiomer (86% *ee*) using the same cofactor.[184] However, diamide is not a universal solution since it is not compatible with all transition-metal based catalysts. Moreover, CFEs-based screening is challenging.

2.3.2 Periplasmic Expression of Sav

In an effort to move away from the highly oxidative environment in the cytosol, and in prospect of increasing the screening throughput, a periplasmic approach was followed. Built on previous studies, T7-tagged Sav was N-terminally fused to the outer membrane protein A (OmpA) leader sequence.[185,186] Using the periplasmic approach where Sav is expressed in the cytosol and secreted into the periplasm of *E. coli* (Sav^{peri}), Jeschek *et al.* optimized a metathase, an ArM catalyzing the ring-closing metathesis (Scheme 4).[187] A 96-well plate assay was devised to screen single clones of *E. coli* expressing Sav in the periplasm. Fourteen positions in close vicinity of the biotinylated Hoveyda-Grubbs second generation catalyst were subjected to saturation mutagenesis and were iteratively screened for improved metathesis activity. Five rounds of evolution yielded a mutant of Sav with 5 single point mutations and about 5-fold increase in catalytic activity.[187]



Scheme 4. Ring closing metathesis of non-fluorescent substrate **31** to umbelliferon **32** by a methathase. The incorporated metallocofactor is a biotinylated second generation Hoveyda-Grubbs catalyst **34**.

Based on the same approach, three reactions catalyzed by ArMs (ring closing metathesis and two deallylation reactions) were recently screened. Variants displaying two mutations each (400 variants in total) were screened using Sav^{peri} and a 96-well plate screening assay. In all three cases, the best mutants significantly improved the catalytic activity of the respective ArM and reached 5- to 20-fold higher TONs than the free cofactor in solution.[188]

2.3.3 Surface-Displayed Sav

One limitation of the periplasmic approach is the uptake of cofactor within the periplasm. For the examples mentioned above, passive diffusion of the synthetic cofactor across the *E. coli* membrane is essential. Surface-display of Sav (Sav^{SD}) was devised as an alternative approach to make the protein more accessible to the cofactor and the substrate. In this case, Sav expressed in the cytosol is exported and displayed on the surface of *E. coli* using an Lpp-OmpA anchor first described in 1992 by Georgiu *et al.*[189] A part of the lipoprotein of *E. coli* (Lpp) was fused to the first five *β*-strands of OmpA and this in turn fused N-terminally to Sav (Figure 17). This leads to the attachment of the Lpp domain into the outer membrane *E. coli* with the five *β*-strands spanning across the membrane and displaying Sav on the outside (Figure 17).

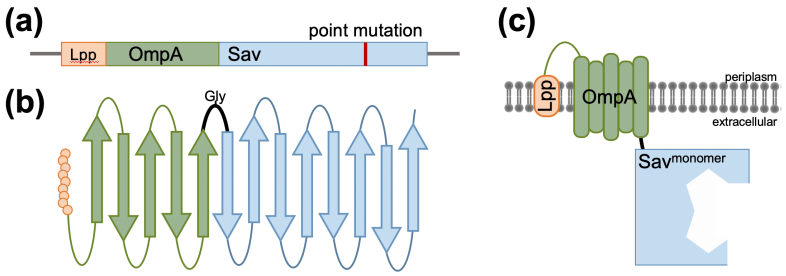
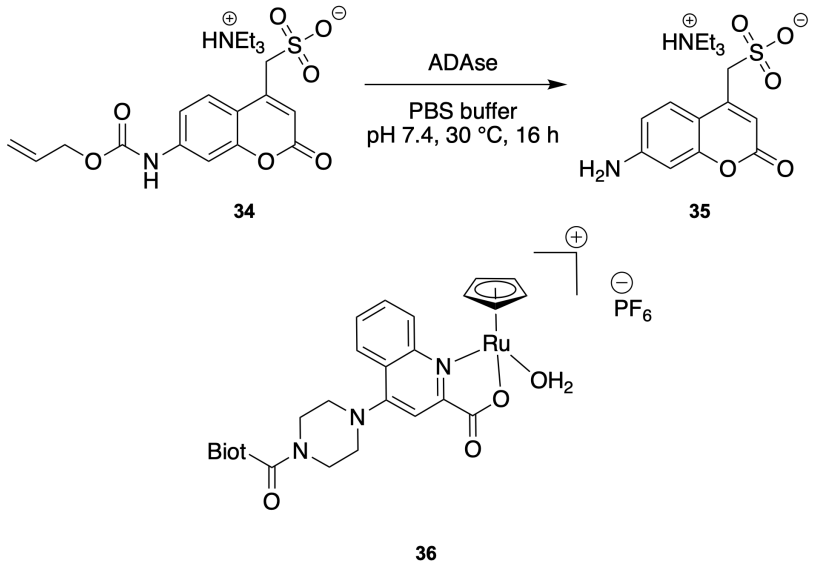


Figure 17. Surface-displayed Sav. a) Schematic representation of the Lpp-OmpA-Sav coding region. **b)** Schematic representation of Lpp-OmpA-Sav secondary structure consisting of nine residues of the Lpp protein (orange), five *β*-sheets of OmpA (green) and eight *β*-sheets of Sav. **c)** Integration of the Lpp-OmpA-Sav construct in the outer membrane of *E. coli* cells. For clarity only one monomer is displayed.

In a recent study, Heinisch and Schwizer *et al.* improved the catalytic activity of an ArM for allylic deallylation (ADAse). Deprotection of an allyl-carbamate-protected coumarin (Scheme 5), served as a model system for the first directed evolution campaign. Libraries at positions K121 and S112 were introduced employing the 22-codon trick and were iteratively screened for catalytic activity using a 96-well plate assay. In the *in vivo* assay, the best mutant (S112Y-K121S) had a 25-fold improvement over WT, whereas, in purified protein, the improvement factor for the best variant (S112M-K121A) was 5.7-fold.[179] This approach was further explored and a catalytic system with a caged inducer for expression of GFP investigated. Disparities between the *in vivo* and the *in vitro* assay, lead to the further investigation of the oligomeric structure of Sav^{SD} (Chapter 3.3).



Scheme 5. Allylic deallylation by an ADAse with the incorporated metallocofactor [CpRu(Biot-Quinoline)(H_2O)] **36** leads to the uncaging of the fluorescent aminocoumarin **35**.

2.3.4 Single Chain Dimer of Sav

Another concern with Sav lies in its homotetrameric nature, which has the consequence that a point mutation will be reflected in all four monomers (Figure 18a). In most reported ArMs based on Sav, the metal center lies at the interface of two adjacent monomers. Residues that were so far shown to have a substantial influence on the catalysis, mainly positions S112 and K121, are at the exact border. Therefore creating an isoform of Sav, where two adjacent monomers could be altered independently from each other was deemed advantageous and a dimeric version of Sav, the single chain dimer (scdSav) was developed.[190] By minimizing the sequence homology, two adjacent Sav monomers are encoded independently and linked via a 26 amino acid linker (Figure 18). Additionally, in Sav^B, H127 was mutated to a cysteine to enforce a disulfide bridge in order to form a tetramer quaternary structure. Moreover, the positions N23, S27 and D128 were mutated to alanine, aspartate and alanine respectively in Sav^B only, to knock out biotin binding and creating the isomer scdSav-mv2, that can only bind two equivalents of biotin (Figure 18b). This isoform thus presents two assets: (i) Analysis of several crystal structures with incorporated biotinylated metal cofactor in homotetrameric Sav, reveals that the metal is located at the interface of two adjacent monomers. Accordingly, mutating the two monomers independently may reveal new variants that were not accessible with the homotetrameric Sav host. (ii) When working with purified Sav, a ratio of biotinylated cofactor to Sav of 2:1 is used to ensure that only one of the two close lying monomers is occupied with a cofactor. However, when working with a library of Sav *in vivo*, the Sav concentration is unknown, making it difficult to impossible to ensure a correct ratio of cofactor to Sav.

A small-scale screening of 33 different scdSav and scdSav-mv2 variants, based on the positions S112 and K121, was conducted and indeed revealed 90% *ee* and >17'700 TON for the best mutant.[190] *In vivo* studies targeting all four critical positions simultaneously (20^4 = 160'000 variants) may bring forth more beneficial quadruple mutants (Chapter 4.5).

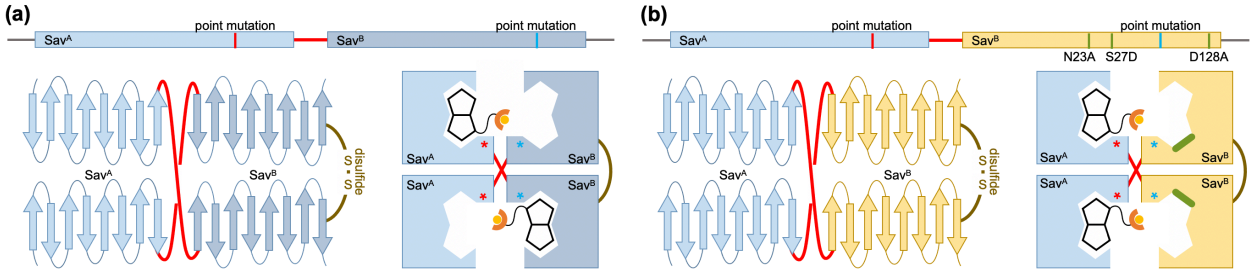


Figure 18. Single-chain dimeric Sav. a) Schematic representation of the scdSav. The sequence homology of monomer A (Sav^A, blue) and monomer B (Sav^B, dark blue) is reduced and the two monomers are linked via a 26 amino acid linker (red). The disulfide bond ensures the correct formation of the tetramer (dark yellow line). Two point mutations are highlighted (red and blue in Sav^A and Sav^B respectively). **b)** Schematic representation of the scdSav-mv2. Three additional mutations (green) knock out the biotin binding in Sav^B. Secondary structure of the scdSav tetramer and quaternary structure of the scdSav with two incorporated biotinylated metallocofactors. Two point mutations are highlighted (red and blue in Sav^A and Sav^B respectively).

2.3.5 SAV-CHIMERAS

Another major throwback of Sav is, that the active site is relatively exposed to the solution, with the effect, that most metal cofactors are not localized within the vestibule in a tight positioning. This could lead to different disadvantages like metal cofactor poisoning by the solvent, and reduced occupancy in the binding pocket disrupting a tight modulation of the transition states. For this reason, several avenues to shield the binding pocket are being tried. Recently one such approach led to the development of a highly active hydroaminase (HMAse) based on a Sav-SOD chimera (Figure 19). Superoxide dismutase C (sodC) from *M. tuberculosis* has a dimerization domain that spans across two Greek key b-barrels of sodC, which represents an ideal size to connect and shield the biotin binding vestibule of Sav. The dimerization domain was engineered between β-sheets 3 and 4 to yield Sav-SOD. This HMAse based on Sav-SOD showed exquisite regioselectivity and was evolved over three rounds in *E. coli* cell free extract and highly active mutants showed either anti-Markovnikov activity (51 TONs and 96% regioselectivity) or Markovnikov activity (333 TONs and 99% regioselectivity).[191]

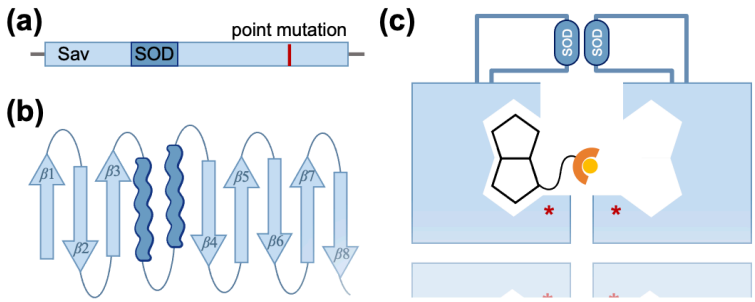


Figure 19. Sav-SOD chimera. a) Schematic representation of the Sav-SOD region. **b)** Schematic representation of the Sav-SOD secondary structure with the dimerization domain (dark blue) inserted between the *β*-sheet 3 and 4. **c)** Schematic representation of the possible orientation of the SOD-cap in respect to Sav.

Apart from the examples mentioned above, several other groups have ventured into this exciting field and about 50 different reactions based on ArMs have been reported so far, including hemoproteins such as P450s,[192] lactococcal multiresistance regulator,[193] human serum albumin,[194] and carbonic anhydrase[195] among others.

But however remarkable the aforementioned results are, they do not yet capitalize on the full potential of directed evolution. Every example, except one, targets only one position per round of evolution, effectively missing out on potentially cooperative beneficial mutations. Combining ArMs and droplet microfluidics, can be a powerful tool for propelling research based on directed evolution forward. Systematic and high-throughput screening of ArMs *in vivo* using double emulsions could allow the screening of a much bigger sequence space, which is to date not feasible. Identifying cooperative effects to improve catalysis or even remodeling whole enzymes to achieve completely new-to-nature reactivities are only two potential examples. Reactions based on ArMs could ultimately provide aqueous and overall environmentally friendly reaction pathways for industrial applications. Additionally, such big data sets could also be used as an input for machine learning applications, to further study active site elasticity, reaction pathways, or even protein folding mechanisms.

2.4 Allylic Substitution Catalyzed by a Ruthenium Pianostool- complex

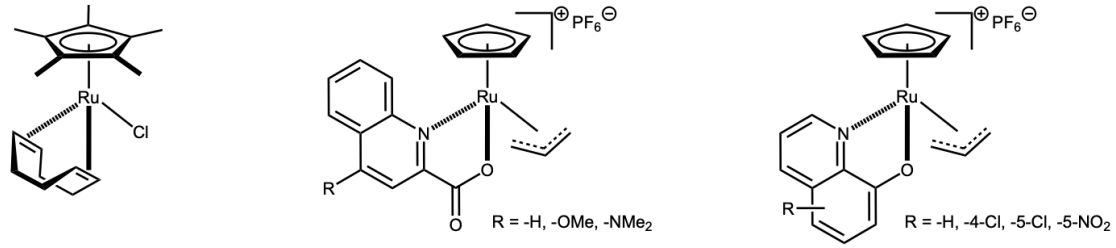
Transition metal-catalyzed reactions are widely used in organic synthesis and offer promising avenues for biorthogonal applications.[196–199] A variety of metal complexes based on palladium,[200] iridium,[201] ruthenium,[202] rhodium,[203] tungsten and molybdenum[204] are known. High modularity, and access to regio- and stereoselectivity are the main assets of such systems and are achieved by the nucleophilic attack onto an η^3-allyl-metal intermediate, where the nature of the

metal-cofactor controls the outcome of the reaction. Ruthenium complexes used for unsymmetrical allyl substitution include complexes based on Cp and Cp*,[205] or derivatives of cyclooctadiene (cod),[206] diimine ligands such as bipyridine (bpy) and diaminobenzidine (dab),[207] amidinate[208] and carbonate[209] are known and are widely used. However, the reaction conditions of these traditional metal catalysts (organic solvent, elevated temperatures, oxygen free atmosphere etc.) are not ideally suitable for biorthogonal applications.

2.4.1 A BIOORTHOGONAL REACTION

The formation of an ArM that can be evolved *in vivo* by directed evolution, needs a metal cofactor that is stable in an aqueous solution, must tolerate oxygen, and is compatible with a cellular environment. Transition metal catalysts which fulfill these requirements include a set of ruthenium-based catalysts introduced by Meggers and coworkers.

As early as 2006, Streu *et al.* reported the first generation of a ruthenium complex capable of uncaging an alloc-protected amine under biological conditions.[210] They reported the ruthenium complex [Cp*Ru(COD)Cl] **37** and observed that it was water and oxygen tolerant, but however needed thiol derivatives as an additive to catalyze the uncaging of the alloc-protected amine. Addition of thiophenol lead to an efficient reaction with 80% conversion in *E. coli* cell lysates. Strikingly, the same catalyst was shown to be active in mammalian cells. Cell viability of HeLa cells was retained, making it a promising starting point. In 2014, Völker *et al.* reported the second generation of the bioorthogonal catalyst [CpRu(QA)(Allyl)]PF_6 (**38**, QA = 2-quinolinecarboxylate)[211] which was first described by Kitamura *et al.* in 2004.[212] Historically, it was used for the allylation of alcohols, the deallylation of allyl ethers, allyl esters and allyl carbonates. Meggers and coworkers applied this ruthenium complex **38** in the deprotection of *O*-allyl carbamates. Deprotection of an *O*-allyl carbamate caged aminocoumarin **44** releases aminocoumarin **45** as the product of the deallylation reaction (Scheme 8). The reactions were performed in water and in the presence of thiols. They further demonstrated the catalytic in-cell activation of a caged drug under biologically relevant conditions, highlighting the applicability of such a system in a biological setting. In 2017, Meggers *et al*. reported an improved ruthenium complex based on a hydroxyquinolinate ligand [CpRu(HQ)(Allyl)]PF_6 (**39**, HQ = hydroxyquinolinate), which exhibits TONs >300 and also showed high activity in blood serum.[213] Furthermore, activity of these complexes in HeLa cells was observed: a caged doxorubicin derivative was deprotected in HeLa cells, inducing cell apoptosis with an administration of a single dose (1 μM) of the catalyst.



37
[Cp*Ru(COD)Cl]

1st generation

38
[CpRu(QA)(Allyl)]PF_6

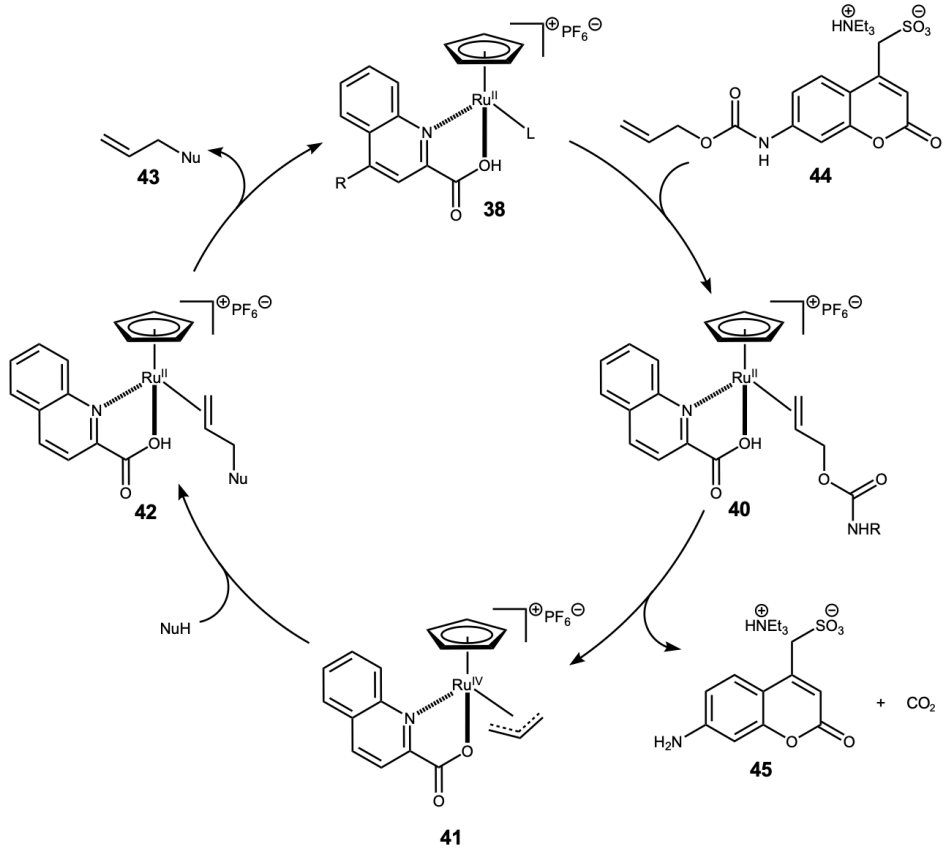
2nd generation

39
[CpRu(HQ)(Allyl)]PF_6

3rd generation

Scheme 6. Three generations of catalysts developed by Meggers.

The ruthenium-catalyzed allylic deprotection of the *O*-allyl carbamate-caged aminocoumarin **44** is assumed to follow the same mechanism as the palladium-catalyzed deallylation reaction (Scheme 7). In a first step, the allyl-substrate **44** coordinates to the Ru^{II}-complex **38**. Decarboxylation and release of the fluorescent leaving group **45** leads to the oxidative addition of the allyl-group, affording a Ru^{IV}-intermediate **41**. Nucleophilic attack on the allyl moiety by a sacrificial nucleophile (NuH, e.g. water or thiols present in the cells) affords the corresponding olefin-Ru^{II}-complex **42**. The weakly-bound olefin is released, thus regenerating the cofactor **38** (Scheme 7).[211]



Scheme 7. Proposed mechanism for the Ru-catalyzed deprotection of *O*-allyl-carbamate-caged aminocoumarin.

These complexes and the deprotection reaction were widely applied by various research groups for *in vivo* catalysis.[214–216] Mascareñas and coworkers initially showed the uncaging of DNA-binding agents such as bisbenzamidine, DAPI and ethidium inside living cells using the first generation [Cp*Ru(COD)Cl] (**37**) catalyst.[217] Rotello and coworkers reported a supramolecular approach using gold nanoparticles, where the [Cp*Ru(COD)Cl] (**37**) complex was embedded inside the particle and the activity was regulated by supramolecular interactions with cucurbit[7]uril.[218] These initial applications highlighted the temporal control of the deprotection reaction, however, taking it a step further and implementing spatial control would allow the localization of the reaction to a desired compartment.

Mascareñas and coworkers included a triphenylphosphonium linker to the second generation catalyst [CpRu(QA-NMe_2)(Allyl)]PF_6 (**38**) which allowed the localization of the catalyst in the mitochondria within HeLa cells. They could show subcellular activity by the release of a protected fluorophore. By using a pyrene-derivatized phosphonium linker, they could track the catalyst's localization.[219]

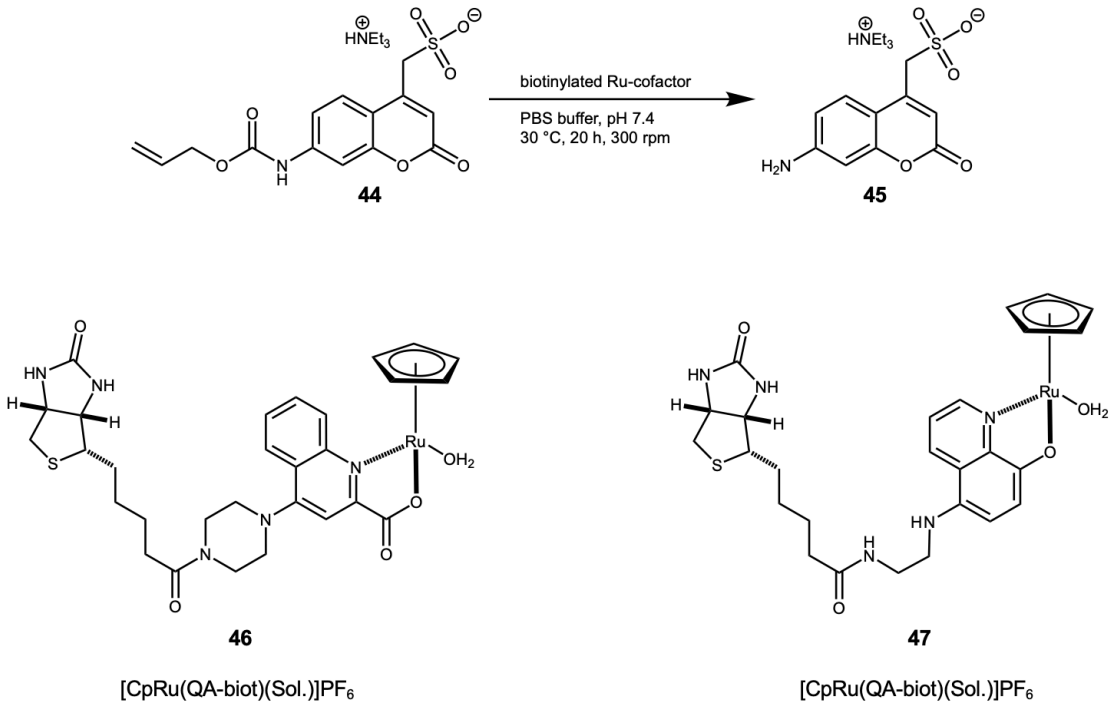
Wender and coworkers developed a real-time monitoring method for intracellular allylcarbamate deprotection using a luciferase reporter-system. They used a cell line with an additional plasmid for luciferase which resulted in bioluminescence, once the deprotected amino-luciferin was produced. However, they could not observe any bioluminescence, if the cells were washed after the catalyst's incubation. Also, decreased catalytic activity was observed if the cells were incubated with the pro-fluorescent amino-luciferase and washed before the addition of the cofactor. They concluded that catalysis might be taking place outside the cells and the substrate might be localized between the intracellular and extracellular space.[220]

More recently, Mayer and coworkers presented a screening strategy that used the 2nd and 3rd generation Meggers' catalysts to generate a non-canonical amino acid that was then integrated within GFP.[221] This approach had two advantages: firstly, it yielded a quantifiable readout to evaluate the efficiency of the catalyst and secondly, it provided a means to assess cell viability. They further demonstrated the rescue of a synthetic *E. coli* auxotroph by using a homogeneous Ru-complex and a heterogeneous Pd-complex that performed a new-to-nature reaction to provide an essential non-canonical amino acid for the survival of the cell.[222]

2.5 Artificial Deallylase (ADAse)

To create ArMs based on the second and third generation complexes **38** and **39**, biotinylated versions of both complexes were synthesized. The biotinylated [CpRu(QA-biot)(Sol.)]PF_6 (cofactor

46) was synthesized according to literature[223] and the biotinylated [CpRu(HQ-biot)(Sol.)]PF_6 (cofactor **47**) was synthesized similarly with a diamine linker (Appendix, Scheme S2). The Sav host protein was reacted with either cofactor **46** or cofactor **47** in a 2:1 ratio (i.e. 2 eq. of cofactor and 1 eq. of tetramer, so that statistically only two of the four biotin binding sites are occupied by the cofactor). The assembled ADAses were then used for the uncaging of the allyl-carbamate protected amino coumarin **44** (Scheme 8). Initial studies of both ADAses revealed the superiority of [CpRu(QA-biot)(Sol.)]PF_6 **46** over [CpRu(HQ-biot)(Sol.)]PF_6 **47** in terms of turnover number after 16 hours. Importantly, we additionally found that: (i) the activity of the ruthenium complex [CpRu(QA)(Allyl)]PF_6 (**38**) is independent of the glutathione concentration (Table 1), and (ii) that Cp performs better than Cp* as a capping ligand. As the newly developed cofactor **47** · Sav did not perform well, cofactor **46** · Sav was selected for the directed evolution campaign. Additional *in vitro* and *in vivo* studies involving various isoforms of Sav and cofactor **46** are summarized in Chapter 4, Figure 37.



Scheme 8. Ruthenium-catalyzed deprotection of allyl-carbamate protected coumarin **44**. Biotinylated ruthenium complexes based on the 2nd generation (**46**) and 3rd generation (**47**) Meggers catalysts.

Table 1. Comparison of the 2nd generation and 3rd generation Meggers cofactors.

Entry	Complex/ArM	Catalyst loading [mol%]	TON	
			30 min	16 h
1[a]	[CpRu(QA-NMe_2)(Allyl)]PF_6 **38**	10	4.1	7.3
2[a]	[CpRu(HQ)(Allyl)]PF_6 **39**	10	2.7	6.4
3[b]	[CpRu(QA-NMe_2)(Allyl)]PF_6 **38**+ GSH	10	4.8	7.6
4[b]	[CpRu(HQ)(Allyl)]PF_6 **39**+ GSH	10	6.6	7.5
5[c]	cofactor **46**	10	0.35	3.4

6[c]	cofactor **47**	10	0.26	0.9
7[c]	cofactor **46** · Sav-wt	10	0.60	4.4
8[c]	cofactor **47** · Sav-wt	10	0.15	0.26
9[d]	cofactor **46**	1	0.7	6.2
10[d]	cofactor **47**	1	0.6	1.2
11[d]	cofactor **46** · Sav-wt	1	1.0	6.6
12[d]	cofactor **47** · Sav-wt	1	0.6	1.1

Reaction conditions: a) substrate **44** (500 μM), ruthenium catalyst in water and DMF (0.5%), 30 °C, air. b) substrate **44** (500 μM), ruthenium catalyst in water and DMF (0.5%), GSH (50 mM), 30 °C, air. c) substrate **44** (500 μM), ruthenium catalyst in water and DMF (0.5%) 30 °C, air. d) substrate **44** (500 μM), ruthenium catalyst in water and DMF (0.5%), GSH (5 mM), 30 °C, air. e) substrate **44** (500 μM), ruthenium catalyst, GSH (5 mM) in PBS (pH 7.4), 30 °C, air. Conversion determined by the measurement of fluorescence intensity of product **45**.

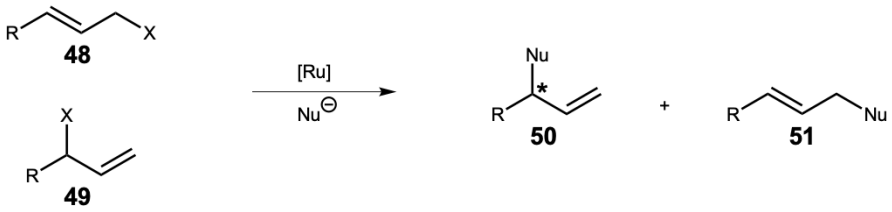
Based on (i) the high activity and *in vivo* compatibility of the biotinylated second-generation Meggers catalyst and (ii) the pioneering work on surface-displayed Sav by Heinisch and Schwizer *et al.*, surface-displayed ADAse was selected as the starting point to optimize its catalytic activity by directed evolution.

Initial steps towards a selection method based on the same [CpRu(QA-biot)(Sol.)]PF_6 **46** · Sav were investigated (Chapter 2.6). Further genetic optimization of surface-displayed ADAse was carried out and the quaternary structure of surface-displayed Sav was elucidated (Chapter 3). And finally, a high-throughput screening assay based on [CpRu(QA-biot)(Sol.)]PF_6 **46** · Sav was developed (Chapter 4).

2.6 ARTIFICIAL CARROLLASE

Besides uncaging reactions, allylic substitution is a powerful tool to design reactions with specific regio- and/or stereoselectivity. This is of high importance, since complex three-dimensional molecules, such as DNA, amino acids and proteins, are ubiquitous in nature and pose a grand challenge. The development of novel drugs or pesticides that can target these structures specifically, makes the synthesis of small molecules, that are equally specific, essential. As mentioned before, several ruthenium-based complexes for allylic substitution have been reported. In an effective asymmetric catalyst, the nature of the allyl-metal intermediate, governs the selectivity of the reaction. In the realm of ruthenium-catalyzed allylic substitutions, primary or secondary allyl-chlorides and allyl-carbonates are common substrates. Ruthenium is known to promote nucleophilic addition to yield branched products (**50**) over linear products (**51**) favoring nucleophilic attack to the most substituted position (Scheme 9).[224,225] However, ruthenium based asymmetric reactions are unusual and C-C bond forming reactions are even rarer. An example describing the kinetic resolution of a racemic allyl-carbonate using a planar-chiral cyclopentadienyl-ruthenium complex

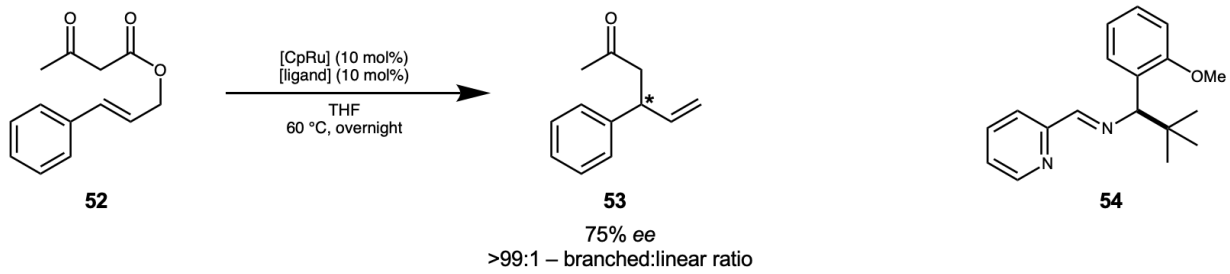
was reported by Takahashi.[226,227] Bruneau and coworkers reported the first Cp*Ru-catalyzed enantioselective reaction – the etherification of cinnamyl chlorides with phenols as nucleophiles.[207]



Scheme 9. A general representation of an allylic substitution reaction leading to branched and linear products.

Expanding this repertoire, Tunge and coworkers described an intramolecular allylic substitution, combining regioselectivity and stereospecificity. In the Carroll rearrangement catalyzed by $[Cp^*RuCl]_4$, the allylated *β*-ketoester **52** undergoes a sigmatropic rearrangement to mainly afford a *γ*,*δ*-unsaturated ketone **53** (Scheme 10).[228,229] Historically the decarboxylative [3,3] sigmatropic intramolecular rearrangement requires harsh conditions such as 140-210 °C and organic solvents, and was first reported by Carroll as early as 1940.[230] The reported reaction was comparatively mild: with CH_2Cl_2 as a solvent and carried out at room temperature. Carroll rearrangement is an interesting reaction to install stereocenters and several synthetic applications are known.[231–233]

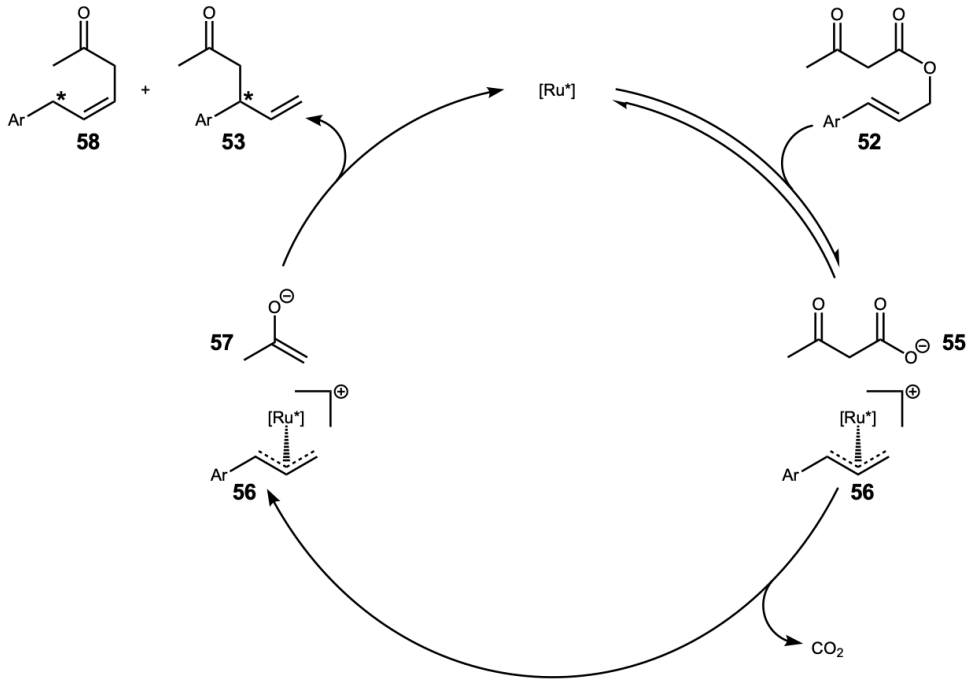
Constant *et al.* improved the intramolecular reaction further by screening various 2,2'-bipyridine (bpy), diimine and pyridine-imine ligands. Instead of Cp*Ru, they used CpRu half-sandwich complexes. Their best complex based on a pyridine-imine ligand (**54**) afforded an 75% *ee* and a branched/linear ratio of >99:1% (Scheme 10).[224] Moreover, they screened different metal precursors and found that if air-stable precursors such as $[CpRu(\eta^6\text{-naphthalene})][PF_6]$ are used, the reaction can be carried out in non-degassed THF and catalyst loadings could be as low as 2 mol%.[225]



Scheme 10. Ruthenium-catalyzed Carroll rearrangement of a *β*-ketoester **52** to a *γ*,*δ*-unsaturated ketone **53**.

The use of CpRu complexes by Lacour and coworkers, and the similarity in the catalytic cycle (Scheme 11), inspired us to use the established ADAse for the rearrangement reaction. The catalytic cycle begins with coordination of the ruthenium cofactor to form an allyl-metal intermediate **56** and

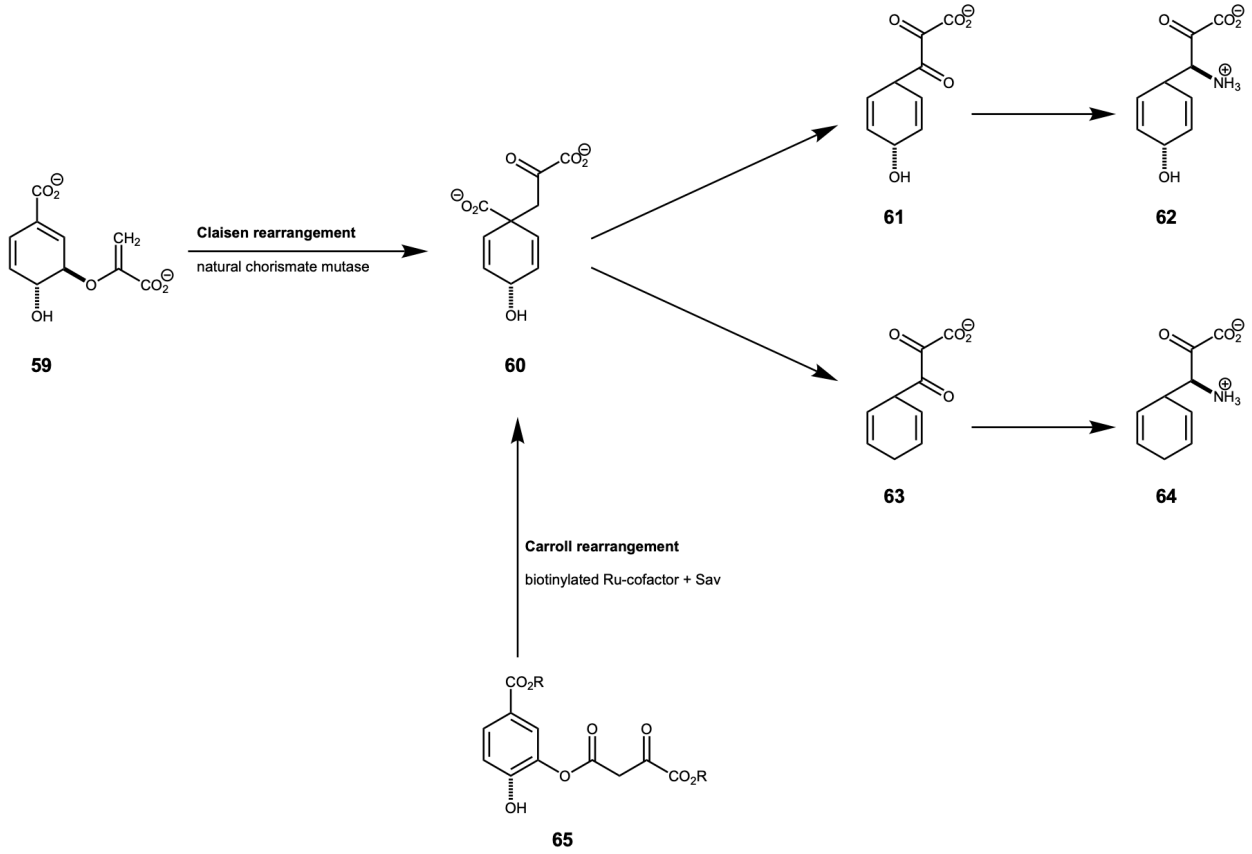
acetoacetate **55**. Decarboxylation of **55** forms the nucleophile **57** which can attack the coordinated allyl either on the less substituted position, forming the linear product **58**, or on the higher substituted position, affording the branched product **53**. In this catalytic cycle, the metal-cofactor governs the regioselectivity (branched or linear). Moreover, in the case of the branched product, the catalyst's chirality influences which enantiomer is formed. While the first step solely relies on the presence of a metal, the selectivity can be modulated by the surrounding ligands. We hypothesized that the second coordination sphere provided by the protein might lead to the enantioenriched products.



Scheme 11. Proposed mechanism for the Ru-catalyzed Carroll rearrangement.[225]

The prospect of directed evolution of a carbon-carbon bond forming reaction using an ArM and the catalyst-controlled regio- and stereoselectivity, were incentives to scrutinize the Carroll rearrangement. Moreover, the prospect of implementing an *in vivo* selection assay, where ArMs would produce an essential intermediate required for the survival of the *E. coli* was an additional motivation. Prephenate **60** is an intermediary metabolite in the shikimate pathway that leads to phenylalanine **64** and tyrosine **62**.[234,235] In nature, chorismate mutase catalyzes the conversion of chorismate **59** to prephenate **60** *via* a Claisen rearrangement (Scheme 12). Hilvert and coworkers have demonstrated the evolution of such enzymes using a selection pressure: they combined tunable transcription and an enzyme degradation tag to apply a selection pressure by reducing the effective intracellular chorismate mutase concentration.[236] Based on these results, we envisioned an ArM for the Carroll rearrangement of **65** leading to prephenate **60** to substitute said chorismate mutase and create a selection pressure based on ArMs (Scheme 12). This would allow not only screening

assays but also a selection scheme for ArMs. The following chapters will give an overview of the initial experiments towards the development such an artificial Carrollase.



Scheme 12. Shikimate pathway. Naturally, prephenate **60** is synthesized from chorismate **59** via a Claisen rearrangement catalyzed by chorismate mutase. Prephenate is a central metabolite that leads to tyrosine **62** and phenylalanine **64**. The selection strategy envisioned involves a Ru-catalyzed Carroll rearrangement of substrate **64** to afford prephenate **60**.

2.6.1 Screening of Reaction Conditions

Using the expertise on ADAses, the aforementioned [CpRu(QA)(Allyl)]PF_6 (**38**) and [CpRu(HQ)(Allyl)]PF_6 (**39**) complexes, as well as their biotinylated derivatives **46** and **47** were screened for catalysis in aqueous solution (Scheme 10 and Table 2). Catalyst loadings of 1-10 mol% were applied. However, the reactivity was so low, that only 5 and 10 mol% catalyst loadings gave detectable product. Because of the poor solubility of the substrate **52** in water, DMF (10%) was used as a co-solvent. The reaction was carried out in glass vials and at a substrate concentration of 2 mM. After completion of the reaction, the product was extracted by the addition of cyclohexane and analyzed by UPC^2 (Appendix A). As a control, the "standard Lacour conditions" with bipyridine (**61**, bpy) as ligand were always tested in parallel (degassed THF, 60 °C, overnight).

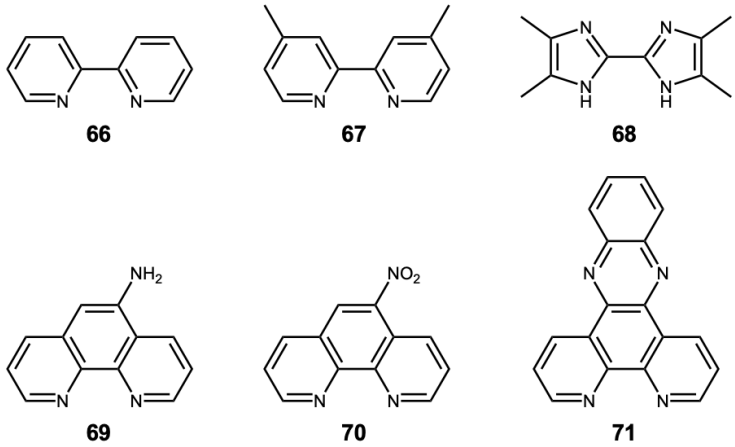
Because of the low activity observed, an array of commercially available diimine ligands (24 in total) was screened in combination with Cp* and Cp ruthenium precursors. As expected, CpRu-

complexes performed much better than Cp*Ru complexes. A summary of the top five ligands can be found in Table 3 (see Appendix A, Table S1 for full list). Phenanthroline-based ligand **70** and ligand **71** performed best with 8.0 and 6.5 TON respectively at a catalyst loading of 1 mol%.

Table 2. Comparison of the 2nd generation and 3rd generation of Meggers cofactors for the Carroll rearrangement.

Entry	Complex/ArM	TON	
		10 mol%	5 mol%
1[a]	bpy **61**	quant.	quant.
2[b]	bpy **61**	1	1
3[b]	[CpRu(QA-NMe_2)(Allyl)]PF_6 (**38**)	1.2	1.5
4[c]	[CpRu(QA-NMe_2)(Allyl)]PF_6 (**38**)+ GSH	1.3	1.3
5[b]	[CpRu(HQ)(Allyl)]PF_6 (**39**)	0.07	-
6[c]	[CpRu(HQ)(Allyl)]PF_6 (**39**) + GSH	0.09	-
7[b]	Biot-**46**	0.12	0.12
8[b]	Biot-**47**	0.03	-

Reaction conditions: a) "standard Lacour conditions" = substrate **52** (2mM), CpRu-bpy catalyst, THF, 60 °C. b) substrate **52** (2 mM), ruthenium catalyst, GSH (10 mM), PBS + DMF (10%), 37 °C, 17 h. c) substrate **52** (2 mM), ruthenium catalyst, GSH (10 mM), PBS + DMF (10%), 37 °C, 17 h.



Scheme 13. Ligands based on bpy and phenanthroline with *N*,*N*-diamine motifs.

Table 3. Selected examples of diimine ligand screening. (see Appendix A, Table S1 for full screening)

Entry	Complex	TON		
		10 mol%	5 mol%	1 mol%
1	Cofactor **46**	0.12	0.12	-
2	Cofactor **47**	0.03	-	-
3	CpRu-**66**			
4	CpRu-**70**	3.5	4.3	8.0
5	CpRu-**71**	2.4	3.7	6.5
6	CpRu-**69**	2.7	2.8	5.5
7	CpRu-**67**	1.7	2.1	1.5
8	CpRu-**68**	1.5	1.4	1.9

Reaction conditions: substrate **52** (2 mM), CpRu-**XX** catalyst, PBS + DMF (10%), 37 °C, 17 h.

To improve the reactions conditions, the addition of the surfactant TPGS-750-M to the mixture, elevated temperatures, and different pHs were evaluated. The best conversions were obtained at pH 7.5 and are summarized in Table 4 (see Appendix A, Figure S2 for other pHs).

Table 4. Improvement of the reaction conditions by the addition of surfactant.

Entry	Complex/ArM	TON	
		37 °C	60 °C
1[a]	bpy **66**	1.6	0.93
2[b]	bpy **66**	0.73	0.90
3[a]	CpRu-**70**	2.9	2.8
4[b]	CpRu-**70**	1.1	1.3
5[a]	cofactor **46**	0.15	-
6[b]	cofactor **46**	0.15	-
7[a]	cofactor **46** · Sav-wt	0.28	0.10
8[b]	cofactor **46** · Sav-wt	0.15	0.06

Reaction conditions: a) substrate **52** (2 mM), ruthenium catalyst (200 µM), Sav (400 µM), water + DMF (10%), pH 7.5, 37 °C, 17 h. b) substrate **52** (2 mM), ruthenium catalyst (200 µM), Sav (400 µM), surfactant TPGS-750-M, water + DMF (10%), pH 7.5, 37 °C, 17 h.

2.6.2 Initial Screening of Sav Variants

In parallel, the biotinylated complexes were tested in the presence of purified protein. 34 available variants with single point mutations at position S112 (11 variants), K121 (13 variants) or L124 (8 variants) were screened. Although the reactivity was rather low, differences between the variants were apparent with Sav-K121M (Table 5) performing the best with ~0.6 TON using cofactor **46**. For the mutants with the highest activity, detectable enantioselectivities could be observed. However, the signals obtained by UPC^2 measurements were so low (Appendix A, Figure S3), that these cannot be used to reliably determine an *ee*.

Table 5. *In vitro* screening of a small Sav library.

Entry	Sav variant	catalyst loading [mol%]	cofactor 46 TON	cofactor 47 TON
1	Sav-wt	10	0.09	0.08
2	K121M	10	0.63	0.13
3	K121V	10	0.31	0.24
4	K121I	10	0.29	0.08
5	K121W	10	0.28	0.17
6	K121A	10	0.21	0.25
7	L124G	10	0.17	0.23

Reaction conditions: substrate **52** (2 mM), CpRu-biot-**46** catalyst (200 µM), Sav (400 µM), water + DMF (10%), 37 °C, 17 h.

From these results it was concluded that the current version of the ADAse using biotinylated complexes by Meggers are not viable catalysts for the Carroll rearrangement. Promising phenanthroline-based ligands (**65** and **66**) were found that could potentially be derivatized with biotin to achieve incorporation into Sav. However initial synthesis approaches were not successful yet.

2.7 REFERENCES

[152] V. Sabatino, J. G. Rebelein, T. R. Ward, *J. Am. Chem. Soc* **2019**, *141*, 24–30.

[156] U. T. Bornscheuer, G. W. Huisman, R. J. Kazlauskas, S. Lutz, J. C. Moore, K. Robins, *Nature* **2012**, *485*, 185–194.

[157] S. Wu, R. Snajdrova, J. C. Moore, K. Baldenius, U. T. Bornscheuer, *Angew. Chemie - Int. Ed.* **2021**, *60*, 88–119.

[158] M. T. Reetz, P. Soni, L. Fernández, Y. Gumulya, J. D. Carballeira, *Chem. Commun.* **2010**, *46*, 8657–8658.

[159] A. D. Liang, J. Serrano-Plana, R. L. Peterson, T. R. Ward, *Acc. Chem. Res.* **2019**, *52*, 585–595.

[160] F. Schwizer, Y. Okamoto, T. Heinisch, Y. Gu, M. M. Pellizzoni, V. Lebrun, R. Reuter, V. Köhler, J. C. Lewis, T. R. Ward, *Chem. Rev.* **2018**, *118*, 142–231.

[161] C. M. Thomas, T. R. Ward, *Chem. Soc. Rev.* **2005**, *34*, 337–346.

[162] A. Radzicka, R. Wolfenden, *Science.* **1995**, *267*, 90–93.

[163] L. Pauling, **1948**, 707–709.

[164] S. J. Benkovic, S. Hammes-Schiffer, *Science.* **2003**, *301*, 1196–1202.

[165] V. L. Schramm, *Chem. Rev.* **2018**, *118*, 11194–11258.

[166] R. A. Sheldon, D. Brady, *ChemSusChem* **2019**, *12*, 2859–2881.

[167] J. Bos, G. Roelfes, *Curr. Opin. Chem. Biol.* **2014**, *19*, 135–143.

[168] J. C. Lewis, **2013**, DOI 10.1021/cs400806a.

[169] F. Schwizer, Y. Okamoto, T. Heinisch, Y. Gu, M. M. Pellizzoni, V. Lebrun, R. Reuter, V. Köhler, J. C. Lewis, T. R. Ward, *Chem. Rev.* **2018**, *118*, 142–231.

[170] B. Kazuo Yamamura, E. Thomas Kaiser, **1976**, *11*, 39–75.

[171] M. E. Wilson, G. M. Whitesides, *J. Am. Chem. Soc.* **1978**, *100*, 306–307.

[172] F. Liu, J. Z. H. Zhang, Y. Mei, *Sci. Rep.* **2016**, *6*, 1–11.

[173] Y. Hiller, E. A. Bayer, M. Wilchek, *Biochem. J* **1991**, *278*, 573–585.

[174] J. A. E. Määttä, Y. Eisenberg-Domovich, H. R. Nordlund, R. Hayouka, M. S. Kulomaa, O. Livnah, V. P. Hytönen, *Biotechnol. Bioeng.* **2011**, *108*, 481–490.

[175] T. R. Ward, *Acc. Chem. Res.* **2011**, *44*, 47–57.

[176] C. M. Dundas, D. Demonte, S. Park, *Appl. Microbiol. Biotechnol.* **2013**, *97*, 9343–9353.

[177] Y. Okamoto, V. Köhler, T. R. Ward, *J. Am. Chem. Soc.* **2016**, *138*, 5781–5784.

[178] A. Chatterjee, H. Mallin, J. Klehr, J. Vallapurackal, A. D. Finke, L. Vera, M. Marsh, T. R. Ward, *Chem. Sci.* **2016**, *7*, 673–677.

[179] T. Heinisch, F. Schwizer, B. Garabedian, E. Csibra, M. Jeschek, J. Vallapurackal, V. B. Pinheiro, P. Marlière, S. Panke, T. R. Ward, *Chem. Sci.* **2018**, *9*, 5383–5388.

[180] J. Serrano-Plana, C. Rumo, J. G. Rebelein, R. L. Peterson, M. Barnet, T. R. Ward, *J. Am. Chem. Soc.* **2020**, *142*, 10617–10623.

[181] H. Renata, Z. J. Wang, F. H. Arnold, *Angew. Chemie Int. Ed.* **2015**, *54*, 3351–3367.

[182] T. Quinto, F. Schwizer, J. M. Zimbron, A. Morina, V. Köhler, T. R. Ward, *ChemCatChem* **2014**, *6*, 1010–1014.

[183] Y. M. Wilson, M. Dürrenberger, E. S. Nogueira, T. R. Ward, *J. Am. Chem. Soc.* **2014**, *136*, 8928–8932.

[184] M. Hestericová, T. Heinisch, L. Alonso-Cotchico, J.-D. Maréchal, P. Vidossich, T. R. Ward, *Angew. Chemie Int. Ed.* **2018**, *57*, 1863–1868.

[185] W. J. Song, F. A. Tezcan, *Science.* **2014**, *346*, 1525–1528.

[186] J. Faraone-Mennella, F. A. Tezcan, H. B. Gray, J. R. Winkler, *Biochemistry* **2006**, *45*, 10504–10511.

[187] M. Jeschek, R. Reuter, T. Heinisch, C. Trindler, J. Klehr, S. Panke, T. R. Ward, *Nature* **2016**, *537*, 661–665.

[188] T. Vornholt, F. Christoffel, M. M. Pellizzoni, S. Panke, T. R. Ward, M. Jeschek, *Sci. Adv.* **2021**, *7*, 4208–4230.

[189] C. Stathopoulos, G. Georgiou, C. F. Earhart, *Appl. Microbiol. Biotechnol.* **1996**, *45*, 112–119.

[190] S. Wu, Y. Zhou, J. G. Rebelein, M. Kuhn, H. Mallin, J. Zhao, N. V. Igareta, T. R. Ward, *J. Am. Chem. Soc.* **2019**, *141*, 15869–15878.

[191] F. Christoffel, N. V Igareta, M. M. Pellizzoni, L. Tiessler-Sala, B. Lozhkin, D C. Spiess, A. Lledose, J.-D. Maréchal, R. L. Peterson, T. R. Ward, Nature Catlysis, accepted.

[192] K. Oohora, A. Onoda, T. Hayashi, *Acc. Chem. Res.* **2019**, *52*, 945–954.

[193] G. Roelfes, *Acc. Chem. Res.* **2019**, *52*, 545–556.

[194] S. Eda, I. Nasibullin, K. Vong, N. Kudo, M. Yoshida, A. Kurbangalieva, K. Tanaka, *Nat. Catal.* **2019**, *2*, 780–792.

[195] J. G. Rebelein, Y. Cotelle, B. Garabedian, T. R. Ward, *ACS Catal.* **2019**, *9*, 4173–4178.

[196] O. Pàmies, J. Margalef, S. Cañellas, J. James, E. Judge, P. J. Guiry, C. Moberg, J. E. Bäckvall, A. Pfaltz, M. A. Pericàs, M. Diéguez, *Chem. Rev.* **2021**, *121*, 4505.

[197] N. A. Butt, W. Zhang, *Chem. Soc. Rev.* **2015**, *44*, 7929–7967.

[198] B. M. Trost, *J. Org. Chem.* **2004**, *69*, 5813–5837.

[199] B. Sundararaju, M. Achard, C. Bruneau, *Chem. Soc. Rev* **2012**, *41*, 4467–4483.

[200] C. Markert, M. Neuburger, K. Kulicke, M. Meuwly, A. Pfaltz, *Angew. Chemie - Int. Ed.* **2007**, *46*, 5892–5895.

[201] J. F. Hartwig, L. M. Stanley, *Acc. Chem. Res.* **2010**, *43*, 1461–1475.

[202] C. Bruneau, M. Achard, *Coord. Chem. Rev.* **2012**, *256*, 525–536.

[203] P. A. Evans, J. D. Nelson, *J. Am. Chem. Soc.* **1998**, *120*, 5581–5582.

[204] A. V. Malkov, I. R. Baxendale, D. Dvořák, D. J. Mansfield, P. Kočovský, *J. Org. Chem.* **1999**, *64*, 2737–2750.

[205] B. M. Trost, P. L. Fraisse, Z. T. Ball, P. C. Ph, *Angew. Chem. Int. Ed* **2002**, *41*, 1059–1061.

[206] T. Kondo, Y. Morisaki, S. Y. Uenoyama, K. Wada, T. A. Mitsudo, *J. Am. Chem. Soc.* **1999**, *121*, 8657–8658.

[207] M. D. Mbaye, J.-L. Renaud, B. Demerseman, C. Bruneau, *J. Organomet. Chem.* **2005**, *690*, 2149–2158.

[208] H. Nagashima, H. Kondo, T. Hayashida, Y. Yamaguchi, M. Gondo, S. Masuda, K. Miyazaki, K. Matsubara, K. Kirchner, in *Coord. Chem. Rev.*, Elsevier, **2003**, pp. 177–190.

[209] I. Fernández, R. Hermatschweiler, F. Breher, P. S. Pregosin, L. F. Veiros, M. J. Calhorda, *Angew. Chemie - Int. Ed.* **2006**, *45*, 6386–6391.

[210] C. Streu, E. Meggers, *Angew. Chemie - Int. Ed.* **2006**, *45*, 5645–5648.

[211] T. Völker, F. Dempwolff, P. L. Graumann, E. Meggers, *Angew. Chem. Int. Ed. Engl.* **2014**, *53*, 10536–10540.

[212] H. Saburi, S. Tanaka, M. Kitamura, *Angew. Chemie - Int. Ed.* **2005**, *44*, 1730–1732.

[213] T. Völker, E. Meggers, *ChemBioChem* **2017**, *18*, 1083–1086.

[214] Y. Bai, J. Chen, S. C. Zimmerman, *Chem. Soc. Rev* **2018**, *47*, 1811.

[215] P. Destito, C. Vidal, F. López, J. L. Mascareñas, *Chem. - A Eur. J.* **2021**, *27*, 4789–4816.

[216] M. O. N. van de L'Isle, M. C. Ortega-Liebana, A. Unciti-Broceta, *Curr. Opin. Chem. Biol.* **2021**, *61*, 32–42.

[217] M. I. Sánchez, C. Penas, M. E. Vázquez, J. L. Mascareñas, *Chem. Sci.* **2014**, *5*, 1901–1907.

[218] G. Y. Tonga, Y. Jeong, B. Duncan, T. Mizuhara, R. Mout, R. Das, S. T. Kim, Y. C. Yeh, B. Yan, S. Hou, V. M. Rotello, *Nat. Chem.* **2015**, *7*, 597–603.

[219] M. Tomás-Gamasa, M. Martínez-Calvo, J. R. Couceiro, J. L. Mascarenãs, *Nat. Commun.* **2016**, *7*, 1–10.

[220] H. T. Hsu, B. M. Trantow, R. M. Waymouth, P. A. Wender, *Bioconjug. Chem.* **2016**, *27*, 376–382.

[221] R. Rubini, I. Ivanov, C. Mayer, *Chem. - A Eur. J.* **2019**, *25*, 16017–16021.

[222] R. Rubini, C. Mayer, *ACS Chem. Biol.* **2020**, *15*, 3093–3098.

[223] T. Heinisch, F. Schwizer, B. Garabedian, E. Csibra, M. Jeschek, J. Vallapurackal, V. B. Pinheiro, P. Marlière, S. Panke, T. R. Ward, *Chem. Sci.* **2018**, *9*, 5383–5388.

[224] S. Constant, S. Tortoioli, J. Müller, J. Lacour, *Angew. Chem. Int. Ed* **2007**, *46*, 2082–2085.

[225] D. Linder, F. Buron, S. Constant, J. Lacour, *European J. Org. Chem.* **2008**, 5778–5785.

[226] Y. Matsushima, K. Onitsuka, T. Kondo, T. Mitsudo, S. Takahashi, *Org. Synth. Mater. Sci.* **2000**, *33*, 37.

[227] K. Onitsuka, Y. Matsushima, S. Takahashi, *Organometallics* **2005**, *24*, 6472–6474.

[228] E. C. Burger, J. A. Tunge, *Org. Lett.* **2004**, *6*, 2603–2605.

[229] E. C. Burger, J. A. Tunge, *Chem. Commun.* **2005**, 2835–2837.

[230] M. F. Carroll, *J. Chem. Soc.* **1940**, 704–706.

[231] M. Defosseux, N. Blanchard, C. Meyer, J. Cossy, *Org. Lett.* **2003**, *5*, 4037–4040.

[232] M. E. Jung, B. A. Duclos, *Tetrahedron Lett.* **2004**, *45*, 107–109.

[233] W. Bonrath, T. Netscher, *Appl. Catal. A Gen.* **2005**, *280*, 55–73.

[234] W. P. Russ, M. Figliuzzi, C. Stocker, P. Barrat-Charlaix, M. Socolich, P. Kast, D. Hilvert, R. Monasson, S. Cocco, M. Weigt, R. Ranganathan, *Science.* **2020**, *369*, 440–445.

[235] P. R. Andrews, G. D. Smith, I. G. Young, *Biochemistry* **1973**, *12*, 3492–3498.

[236] M. Neuenschwander, M. Butz, C. Heintz, P. Kast, D. Hilvert, *Nat. Biotechnol.* **2007**, *25*, 1145–1147.

3

SURFACE DISPLAY OF STREPTAVIDIN

3.1 INTRODUCTION

Surface display of proteins on *E. coli* is a versatile technique that was reported as early as 1986.[237,238] Freudl *et al.* used the outer membrane protein A (OmpA) to display small peptides by inserting non-OmpA (protease cleavable) sequences in between OmpA sequences. Intact cells that had the protein expressed, were treated with protease and analyzed by SDS-PAGE, demonstrated the surface display of these small peptides.[238] The system was then expanded to bigger enzymes and it could be shown that OmpA fused to a lipoprotein (Lpp) was an efficient hybrid protein that could serve as a vehicle to export virtually any protein to the surface of *E. coli*.[189,239]

The main advantage of this technique is the better accessibility of the protein of interest. In contrast to proteins that are located inside the cell, surface-displayed proteins do not need to be tediously extracted and purified, which has a significant time/cost benefit. Many microorganisms such as bacteria and yeast but also viruses and even phages have been reported to display peptides or enzymes on their surface with applications in live cell vaccine production, peptide library screening, surface tagging and whole cell biocatalysis. Especially in the field of biocatalysis, surface-display is an advantage as the enzymes are readily accessible to the reagents, and at the same time, are compartmentalized from the cellular matrix, which may have deleterious effects on the reagents and cofactors.[240–242]

To display streptavidin (Sav) on the surface of *E. coli*, systems based on the autotransporter AIDA or the Lpp-OmpA systems are reported.[243,244] The system used here, and first reported by Georgiou *et al.* is composed of the Lpp signal peptide and the first part of the *E. coli* lipoprotein

(residues 1-9) fused to the outer membrane protein A (residues 46-159). T7-tagged Sav is attached N-terminally to Lpp-OmpA via a glycine linker (Figure 20). The resulting Lpp-OmpA-Sav system is introduced into a pBAD33 vector, where expression of Sav can be induced by the addition of L-arabinose. Lpp anchors to the inner-side of the outer membrane of *E. coli* and the five *β*-strands of OmpA span across the membrane to project Sav to the outside of the membrane.[244]

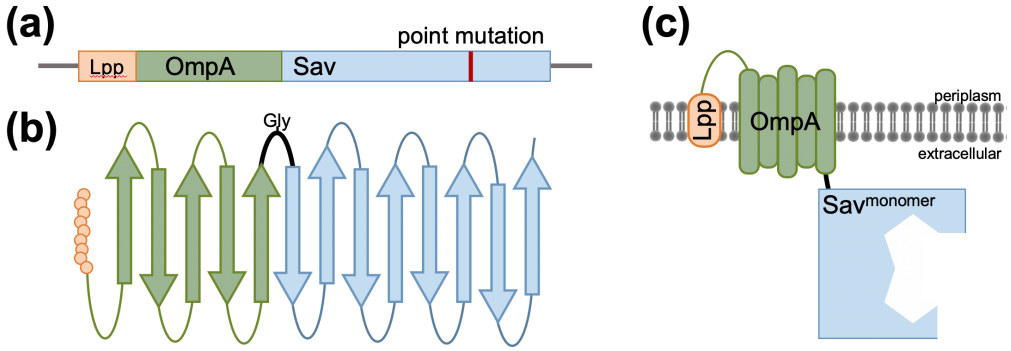


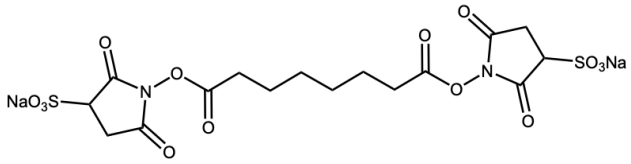
Figure 20. Surface-displayed Sav. a) Schematic representation of the Lpp-OmpA-Sav coding region. **b)** Schematic representation of Lpp-OmpA-Sav secondary structure consisting of nine residues of the Lpp protein (orange), five *β*-sheets of OmpA (green) and eight *β*-sheets of Sav. **c)** Integration of the Lpp-OmpA-Sav construct in the outer membrane of *E. coli* cells. For clarity only one monomer is displayed.

The effective surface-display of Sav (Sav^{SD}) was analyzed by two means: by the addition of a biotinylated Atto-dye (Appendix B, Figure S14) and by the staining of the cells using a fluorescent Sav antibody.[244] Although the biotinylated dye as well as the Sav antibody confirm the presence of Sav^{SD}, the quaternary structure of Sav remains unclear. To form one homotetrameric unit of Sav, four individual units of Lpp-OmpA-Sav need to be in close proximity on the surface of *E. coli*. Since only one glycine is engineered as a linker between the truncated OmpA and Sav, and because of the relatively large size of Sav, the formation of a tetramer is not trivial and needs to be evaluated more carefully. Indeed, catalytic activities of different mutants observed *in vivo* using Sav^{SD} vary to the activities observed *in vitro* of the same mutants (see Chapter 3.3). As mentioned above, the active site of many ArMs based on Sav lies at the interface of two monomers, with the effect that known beneficial mutations are located in the adjacent monomer. Therefore, the correct assembly of a tetramer, or at least a dimer, might be essential for efficient catalysis. To examine these disparities and to unambiguously determine the oligomeric state of Sav^{SD}, a crosslinking experiment involving live cells was envisioned.

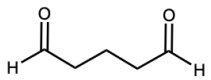
3.2 CROSSLINKING EXPERIMENTS

3.2.1 VALIDATION OF THE CROSSLINKING STRATEGY USING PURIFIED SAV

Initial experiments were carried out using purified wild-type Sav (pp-Sav). Two different crosslinkers, glutaraldehyde **73** and bis(sulfosuccinimidyl) suberate (**72**, BS^3) were tested as crosslinkers (Scheme 14).[245,246] Both reagents are known to target primarily lysines on the surface of the proteins, thereby fixing the quaternary structure. pp-Sav was treated with glutaraldehyde and BS^3 either prior or following denaturation (Figure 21). Denaturation of Sav is not trivial and was achieved by boiling the samples in SDS sample buffer (2% SDS) at > 95 °C for a minimum of 30 min.[247] This step was necessary to validate the crosslinking efficiency. To visualize Sav by UV detection, generally, a biotinylated dye is used. Therefore, initial crosslinking experiments of pp-Sav samples were conducted with and without the addition of the biotinylated dye biotin-4-fluorescein (B4F) (see Appendix B, Figure S13for UV detection).



72



73

Scheme 14. Two crosslinkers used for this study. Bis(sulfosuccinimidyl) suberate **67** and glutaraldehyde **68**.

Experiments following typical crosslinking protocols (1.2 mM crosslinker, 1-5 min incubation on ice) revealed initial crosslinking of pp-WT (Figure 21). Glutaraldehyde led to over-crosslinking, where also denatured samples migrated as tetramers and higher oligomers (Figure 21, lane 6-8). Comparison of the non-crosslinked samples with the crosslinking experiments with BS^3 revealed the following trends:

1. pp-Sav that was not crosslinked migrates as tetramer and higher oligomers, with only faint bands for the dimeric and monomeric structures (Figure 21, lane 1/3). Not crosslinked, but denatured pp-Sav samples mainly migrates as monomer. In addition, these samples also reveal bands for the dimer and for the tetramer structures (Figure 21, lane 2). Addition of the biotinylated dye B4F, renders the denaturation more difficult; accordingly, the tetrameric state is mostly retained (Figure 21, lane 4).
2. pp-Sav that was crosslinked with BS3, migrates as tetramer and higher oligomers (Figure 21, lane 11/13). First crosslinked and then denatured samples, migrate as monomers, dimers, tetramers and higher oligomers (Figure 21, lane 12). Importantly, addition of B4F makes the denaturation virtually impossible (Figure 21, lane 14).

3. pp-Sav that was first denatured, before the addition of B4F, mainly migrates as monomer and partly as dimer (Figure 21, lane 5). pp-Sav that was first denatured, then crosslinked and B4F added, only migrates as monomer (Figure 21, lane 15).

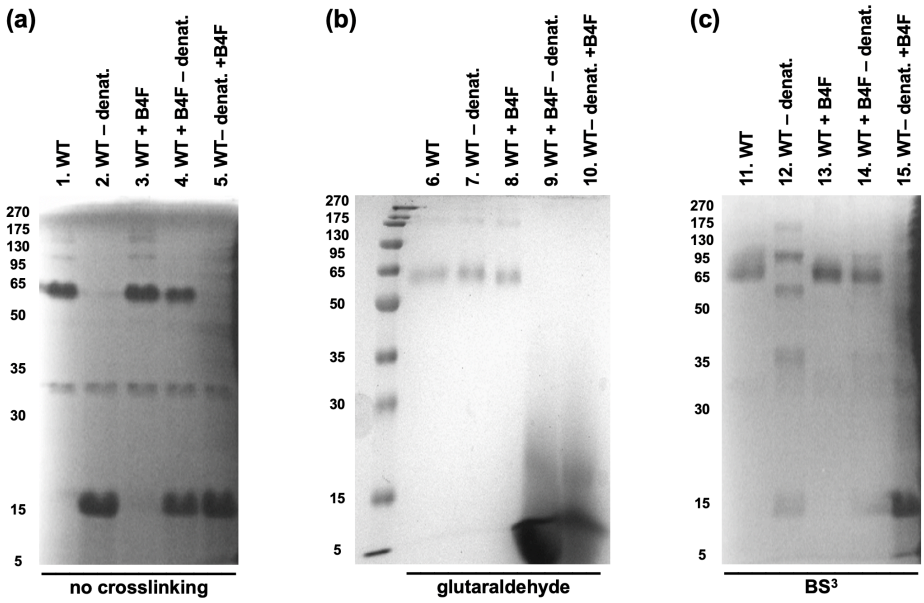


Figure 21. Comparison of different crosslinkers by SDS-PAGE (14% polyacrylamide gel). **a)** Not crosslinked pp-Sav. **b)** Crosslinked with glutaraldehyde. **c)** Crosslinked with BS^3. 1) pp-Sav, 2) denatured after crosslinking, 3) pp-Sav incubated with B4F before crosslinking, 4) pp-Sav incubated with B4F, denatured and then crosslinked, 5) pp-Sav denatured first, then incubated with B4F and finally crosslinked. In the not crosslinked experiment, all samples were prepared the same way, just without the addition of a crosslinker.

In conclusion, BS^3 was deemed more suitable for the following crosslinking experiments. Moreover, since the addition of B4F does not allow the full denaturation of Sav, another way to detect the presence of Sav, a Western blot, was used for the subsequent experiments.

First experiments under the same conditions, seemed to work when only looking at the SDS-PAGE. The crosslinked sample mostly retained its tetrameric structure (Figure 22, lane 4). However, Western blot analysis, unfortunately revealed that the crosslinking was not complete. The denaturation step after crosslinking, that is essential to validate that no aggregates are formed after the crosslinking step, lead to multiple fragments (Figure 22, lane 9). A variety of BS^3 concentrations (1-16 mM) and prolonged incubation times (30 min) at room temperature were evaluated (Figure 22). With a final concentration of 8 mM BS^3 and 30 min incubation at room temperature, the oligomeric state of pp-Sav could be permanently fixed with BS^3 without it being affected by the subsequent denaturation steps.

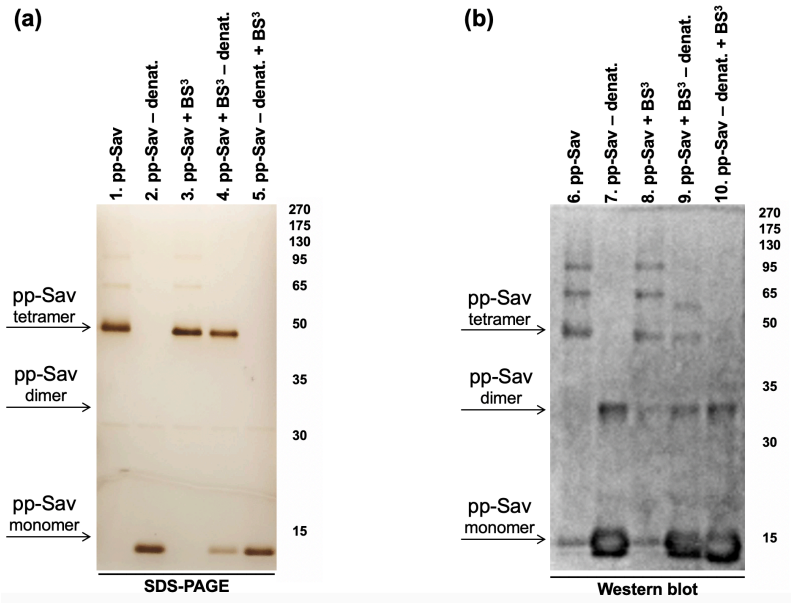


Figure 22. Comparison of an SDS-PAGE to the corresponding Western blot. a) SDS-PAGE (14% polyacrylamide gel, visualized using a silver stain). **b)** Western blot (anti-Sav rabbit polyclonal antibody used at a 1:200 dilution). 1) pp-Sav, 2) denatured pp-Sav, 3) crosslinked pp-Sav, 4) crosslinked pp-Sav, denatured, 5) pp-Sav denatured first, then crosslinked.

3.2.2 Applying the Crosslinking Strategy to Fix the Oligomeric State of Sav^{SD}

To determine the quaternary structure, the main idea was to fix the oligomeric state of surface-displayed Sav (Sav^{SD}) with the conditions developed on live *E. coli* cells, lyse the cells and analyze the sample via SDS-PAGE and Western blot. Two different constructs were analyzed and compared:

In the first approach, Lpp-OmpA-Sav (Sav^{SD}) was expressed and crosslinked. In theory, subsequent lysis of the cells is assumed to yield fragments with the sizes 32 kDa, 64 kDa and 128 kDa for the monomer, dimer and tetramer of Sav^{SD} respectively (Figure 23).

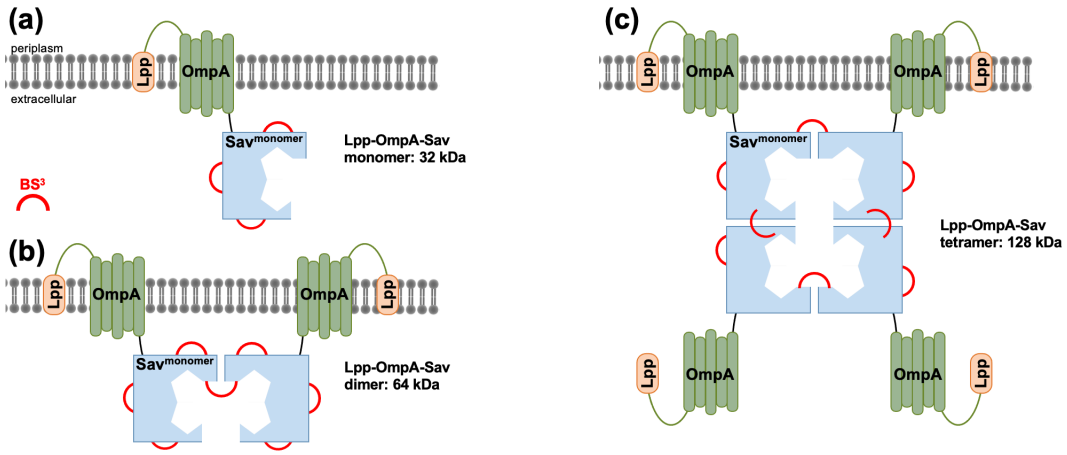


Figure 23. Schematic representation of Sav^{SD} using the Lpp-OmpA display system and the crosslinking approach.
a) Crosslinking with bis(sulfosuccinimidyl) suberate (BS^3), which mainly targets lysines, fixes the oligomeric state of the surface-displayed Sav^{SD}. Subsequent lysis of the cell membrane, releases the crosslinked constructs into solution that are analyzed by SDS-PAGE and western blot. Depending on the surface-displayed quaternary structure of Sav^{SD}, three different sizes, **a)** 32 kDa, **b)** 64 kDa and **c)** 128 kDa can be expected for the monomer, dimer and tetramer of Sav^{SD} respectively.

In the second approach, a construct with an additional Small ubiquitin-related modifier (SUMO)-tag was designed. SUMO proteins are 11 kDa proteins that are mainly fused N-terminally to a target protein to improve its solubility and stability.[248,249] The small protein's structure is closely related to ubiquitin, and shows a *ββαββαβ*-fold, with a long N-terminal tail as determined by NMR.[250] The ubiquitin-like-specific protease 1 (Ulp1), originally found in *S. cerevisiae*, specifically recognizes the tertiary structure of SUMO and cleaves it from the target protein, generating the native protein.[251–253] Based on the expertise and initial studies by Dr. Ryan Peterson in the Ward-lab, who had initially used the SUMO-tag to generate soluble cytosolically expressed streptavidin-superoxide dismutase chimeras (Sav-SOD), a new construct with the SUMO-tag between the OmpA and the Sav moiety was engineered, Lpp-OmpA-SUMO-Sav (Sav^{SUMO}). Adding the SUMO moiety between OmpA and Sav therefore allows the cleavage of the SUMO-tag by Ulp1. Several studies demonstrated the feasibility of such fusion systems for the purification of target proteins such as recombinant FGF21,[254] mCherry[252] or GFP.[255] Thus, it was envisioned that after crosslinking, the SUMO-tag can be cleaved by Ulp1, effectively purifying Sav from cell debris. In this approach, ideally fragment sizes of 16 kDa, 32 kDA and 64 kDa for the monomer, dimer and tetramer of Sav should be observed. If the cleavage is unsuccessful/incomplete the fragments with sizes 43 kDa, 86 kDA and 171 kDa for the monomer, dimer and tetramer of the Sav^{SUMO} constructs will be observed respectively (Figure 24).

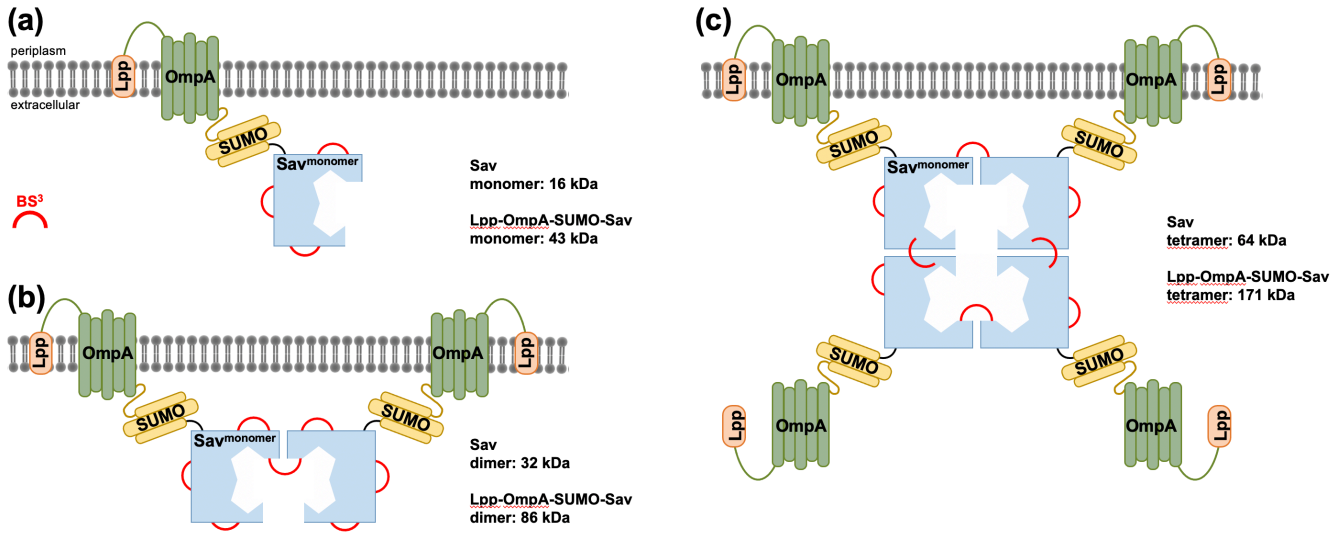


Figure 24. Schematic representation of Sav^{SUMO} using the Lpp-OmpA display system and the crosslinking approach. a) Crosslinking with bis(sulfosuccinimidyl) suberate (BS^3) fixes the oligomeric state of the surface-displayed Sav^{SUMO}. Subsequent proteolytic cleavage with Ulp1 and lysis of the cell membrane, ideally releases crosslinked Sav into the solution that are analyzed by SDS-PAGE and western blot. If the proteolytic cleavage is complete, depending on the surface-displayed quaternary structure of Sav, three different sizes, **a)** 32 kDa, **b)** 64 kDa and **c)** 128 kDa can be expected for the monomer, dimer and tetramer of Sav respectively. If the cleavage is incomplete another three different sizes, **a)** 43 kDa, **b)** 86 kDa and **c)** 171 kDa can be expected for the monomer, dimer and tetramer of Sav^{SUMO} respectively.

The constructs Sav^{SD} and Sav^{SUMO} –as well as the empty vector as negative control– were expressed in *E. coli* at 25 °C for 24 h in shaking flasks. The cells were harvested and the pellets were washed with PBS before cross-linking with BS^3 (8 mM, 30 min, 25°C). The cross-linking was quenched by addition of 1 M Tris-buffer, the cell suspension centrifuged, the supernatant discarded and the cell pellet was resuspended in lysis buffer. The suspension was centrifuged and the supernatant was set aside for analysis. One part of the supernatant was loaded onto the gel (Figure 25, lane 6/9) and one part was denatured (Figure 25, lane 7/10). The remaining cell pellet was resuspended in 8 M urea to ensure the recovery of any additional protein that might have been retained in the cell debris, centrifuged and the supernatant was subjected to analysis after denaturation (Figure 25, lane 8/11).

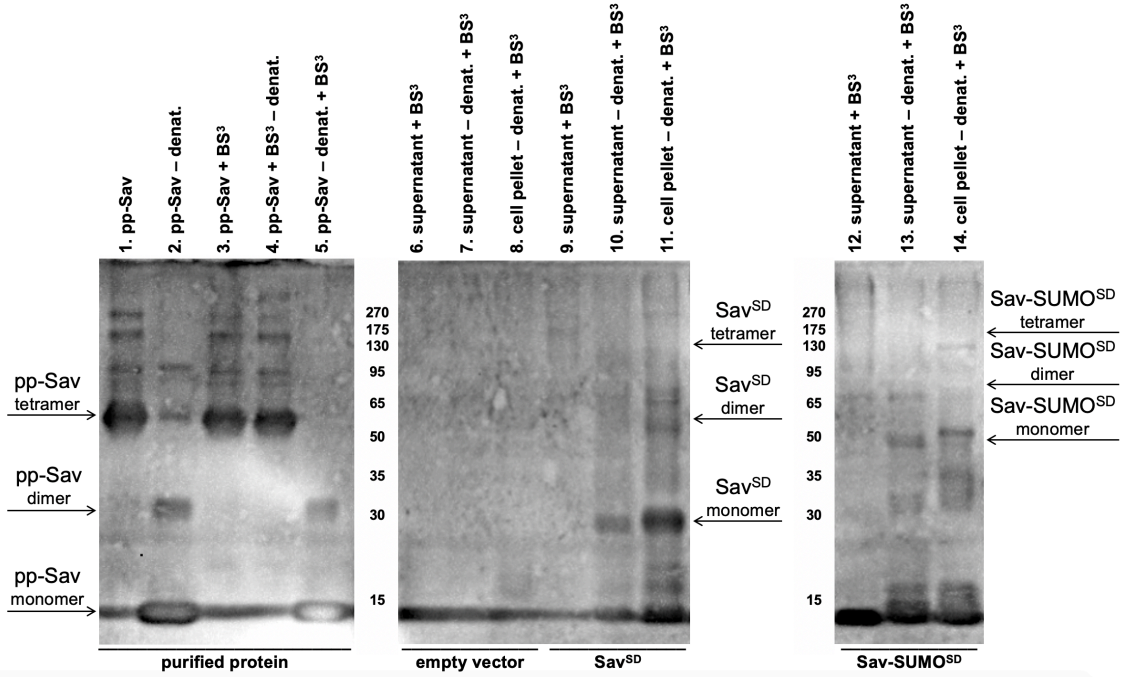


Figure 25. Western blot analysis of the quaternary structure of Sav^{SD}. Crosslinking of Sav^{SD}, followed by SDS-PAGE (Appendix B, Figure S15) and western blot (anti-Sav rabbit polyclonal antibody used at a 1:200 dilution). Except for lane 5, all samples were crosslinked with BS^3 prior to any additional treatment. All denatured samples were treated under the same conditions: SDS sample buffer, 95°C, 30 min. 1) pp-Sav, 2) denatured pp-Sav, 3) crosslinked pp-Sav, 4) crosslinked and then denatured pp-Sav, 5) denatured and then crosslinked pp-Sav, 6) empty vector, supernatant, 7) empty vector, denatured supernatant, 8) empty vector, denatured cell pellet 9) Sav^{SD}, supernatant, 10) Sav^{SD}, denatured supernatant, 11) Sav^{SD}, denatured cell pellet, 12) Sav^{SD}, supernatant, 13) Sav^{SD}, denatured supernatant, 14) Sav^{SD}, denatured cell pellet.

For Sav^{SD}, the electrophoresis results performed on a crosslinked Sav^{SD} sample, strongly suggests that Sav^{SD} is present as a monomer on the surface. Indeed, the denatured samples (Figure 25, lane 10 and 11) display a prominent band at about 32 kDa, which corresponds to the molecular weight of monomeric Sav^{SD} (Lpp-OmpA-Sav = 32 kDA). In contrast, for the non-denatured Sav^{SD} sample (Figure 25, lane 9) faint bands at higher molecular weight are visible. Despite the chaotropic nature of the SDS-PAGE analysis, the remarkable stability of homotetrameric Sav drives its oligomerization (tetramer and higher oligomers), as observed for the pp-Sav (Figure 25,

lane 1-5).[256] For the SavSUMO construct, the electrophoresis results, reveal the same trend. Here, the most prominent band is at about 16 kDa which corresponds to the molecular weight of monomeric Sav (Sav = 16 kDa) (Figure 25, lane 12-14). Moreover, in the denatured samples (Figure 25, lane 13/14), the most prominent signal is at around 50 kDa which could correspond to the molecular weight of monomeric SavSUMO (Lpp-OmpA-SUMO-Sav = 43 kDa).

Based on these qualitative observations, it can be concluded that Lpp-OmpA-Sav is most probably not displayed as a homotetramer on the outer-membrane of *E. coli*. Accordingly, optimization of surface-displayed Sav might generate false positive/negative hits, since only one monomer is present on the surface to form the catalytically active ArM. As mentioned, the active site of Sav lies at the interface of two monomers, and mutations beneficial for catalysis are often located on the adjacent unit. Thus at least the formation of a dimer is essential for the generation of an efficient ArM. In the following chapter, disparities between *in vivo* experiments using Sav^{SD} and *in vitro* experiments using pp-Sav illustrate this phenomenon more clearly.

The following section has been published in
ACS Catal. **2021***, 11,* 1070-510712.
doi: 10.1021/acscatal.1c02405

Directed Evolution of a Surface-Displayed Artificial Allylic Deallylase Relying on a GFP Reporter Protein

Fabian Schwizer[a], Alain Baiyoumy[a], Jaicy Vallapurackal[a], Tillmann Heinisch[a],

Tsvetan Kardashliev[b], Martin Held[b], Sven Panke[b] and Thomas R.Ward[a],*

[a] Department of Chemistry, University of Basel, Mattenstrasse 24a, CH-4058 Basel, Switzerland. E-mail: thomas.ward@unibas.ch

[b] Department of Biosystems Science and Engineering, ETH Zurich, Mattenstrasse 26, CH-4058 Basel, Switzerland.

KEYWORDS: directed evolution · *in vivo* screening · surface-display

3.3.1 ABSTRACT

Artificial metalloenzymes (ArMs) combine characteristics of both homogeneous catalysts and enzymes. Merging abiotic and biotic features allows for the implementation of new-to-nature reactions in living organisms. Here we present the directed evolution of an artificial metalloenzyme based on *E. coli* surface-displayed streptavidin (Sav^{SD} hereafter). Through the binding of a ruthenium-pianostool cofactor to Sav^{SD}, an artificial allylic deallylase (ADAse hereafter) is assembled, which displays catalytic activity towards the deprotection of alloc-protected *m*-hydroxyaniline. The uncaged *p*-aminophenol acts as a gene switch and triggers the over-expression of a fluorescent GFP reporter protein. This straightforward readout of ADAse activity allowed the simultaneous saturation mutagenesis of two amino acid residues in Sav near the ruthenium cofactor, expediting the screening of 2762 individual clones. A 1.7-fold increase of *in vivo* activity was observed for Sav S112T-K121G compared to the wild-type Sav (Sav WT). Finally, the best performing Sav isoforms were purified and tested *in vitro*. For Sav S112M-K121A, a total turnover number of 186 was achieved, corresponding to a 5.9-fold increase *vs.* Sav WT. To analyze the marked difference in activity observed between the surface-displayed and purified ArMs, the oligomeric state of Sav^{SD} was determined. For this purpose, crosslinking experiments of *E. coli* cells over-expressing Sav^{SD} were carried out, followed by SDS-Page and Western blot. The data suggest that Sav^{SD} is most likely displayed as a monomer on the surface of *E. coli.* We hypothesize that the difference between the *in vivo* and *in vitro* screening results may reflect the difference in oligomeric state of Sav^{SD} *vs.* soluble Sav (monomeric *vs.* tetrameric). Accordingly, care should be applied when evolving oligomeric proteins using *E. coli* surface-display.

3.3.2 INTRODUCTION

Artificial metalloenzymes[171] (ArMs) are formed by the incorporation of a synthetic cofactor into a natural host protein.[169,257] Artificial metalloenzymes combine the advantages of synthetic transition metal catalysts (e.g. broad range of reactivities[258–266], large substrate scope, wide choice of metals) with the benefits of enzymes (e.g. high selectivity, high TON, aqueous compatibility, directed evolution, etc.).[156,262,263,265,267–270] Hence, such entities can be optimized by combining both chemical and genetic means, which significantly broadens the range of possible optimization approaches.[271] The effect of the chemical or biological modifications can then be evaluated via a high-throughput screening using various analytical readouts, including fluorescence or UPLC-MS.[272,273]

Building on the pioneering work of Meggers[210,211] we performed the *in vivo* deallylation of an alloc-protected aminocoumarin moiety using an artificial allylic deallylase (hereafter

ADAse).[179] The catalytic system, comprising an air and thiol tolerant ruthenium (II) complex **1**,[211] is anchored within Sav displayed on *E. coli*'s outer-membrane.[179,274] The efficiency of this system was improved *via* iterative saturation mutagenesis of the position S112 and K121.[269,275] The reaction was monitored via the fluorescence of the uncaged aminocoumarin **45** (Figure 26). We know from previous Sav-ArM evolution campaigns, that the iterative site-saturation mutagenesis at positions S112 and K121 often leads to significant improvement in catalytic performance, as these two positions lie in close proximity to the cofactor.[179,271,276] The possibility of expressing a protein at the surface of a cell[264,277–279] dragged our attention since via this process we would avoid many issues encountered via periplasmic or cytoplasmic expression (reductive conditions, lack of accessibility etc.).

In previous studies from Wender, Schwaneberg, Mayer, Meggers and our lab it has been demonstrated the possibility of performing chemical transformation at the surface of living cells[260,261,264] and performing the uncaging of a fluorophore or an inducer in living cells.[179,213,280] More recently, we engineered a gene circuit in the cytoplasm of mammalian cells[281]. Based on these results, we aimed to develop a gene switch based on the regulatory protein DmpR[282–285] to promote the production of a protein of interest and allow us to follow the development of this system via fluorescence. Building on these results, we set out to develop and optimize an ArM whose activity could induce the overexpression of a reporter protein in *E. coli*. With this goal in mind, we selected the [CpRu]-catalyzed deallylation of amines, as pioneered by Meggers. We opted for an *E. coli* surface display of Sav to evolve the ADAse activity, relying on a fluorescent readout caused by the up-regulation of a green fluorescent protein reporter protein (GFP hereafter).

3.3.3 Results and Discussion

To up-regulate the GFP reporter protein in *E. coli*, we selected a DmpR-regulated GFP construct, whereby binding of various aniline derivatives to the DmpR regulator, turns-on the over-expression of GFP, which can be conveniently used as read-out of ADAse activity (Figure 26). As a first step, we tested the induction capacity of various polar benzene derivatives in the presence of reporter cells equipped with the DmpR-regulated GFP plasmid (Figure 27a). Both 2-methylphenol and 3-hydroxyaniline **2** performed best, as reflected by the GFP fluorescence observed in the reporter cells. As allylcarbonates (to protect 2-methylphenol) are more prone to spontaneous hydrolysis than allylcarbamates (to protect 3-hydroxyaniline), we selected the *N*-allylcarbamate-3-hydroxyaniline **3** for further studies. Next, we spiked the DmpR-GFP equipped reporter-cells with the wild-type ADAse consisting of cofactor **46** · Sav-wt (Figure 27b). We were pleased to observe the following trends: i) the protected aniline **74** does not lead to marked GFP-fluorescence in the

reporter cells, ii) the cofactor **46** alone is not very active at deprotecting compound **74**, and iii) cofactor **46** · Sav WT leads to increased GFP-fluorescence (1.8-fold *vs.* cofactor **46** alone). The GFP fluorescence produced in the presence cofactor **46** · Sav WT corresponds to the equivalent of spiking GFP-expressing cells with >100 mM 3-hydroxyaniline **75**. Based on these findings, we engineered a system based on two plasmids: Lpp-OmpA-Sav[274,286] and DmpR-GFP[282,284,285], thus allowing us to screen and optimize the ADAse activity using a single cell. Lpp-OmpA-Sav regulates the expression and translocation of Sav to the surface of *E. coli* cells and is induced by *L*-arabinose. DmpR-GFP expresses a GFP reporter-protein via the binding of an inducer to the DmpR regulator (Figure 26).

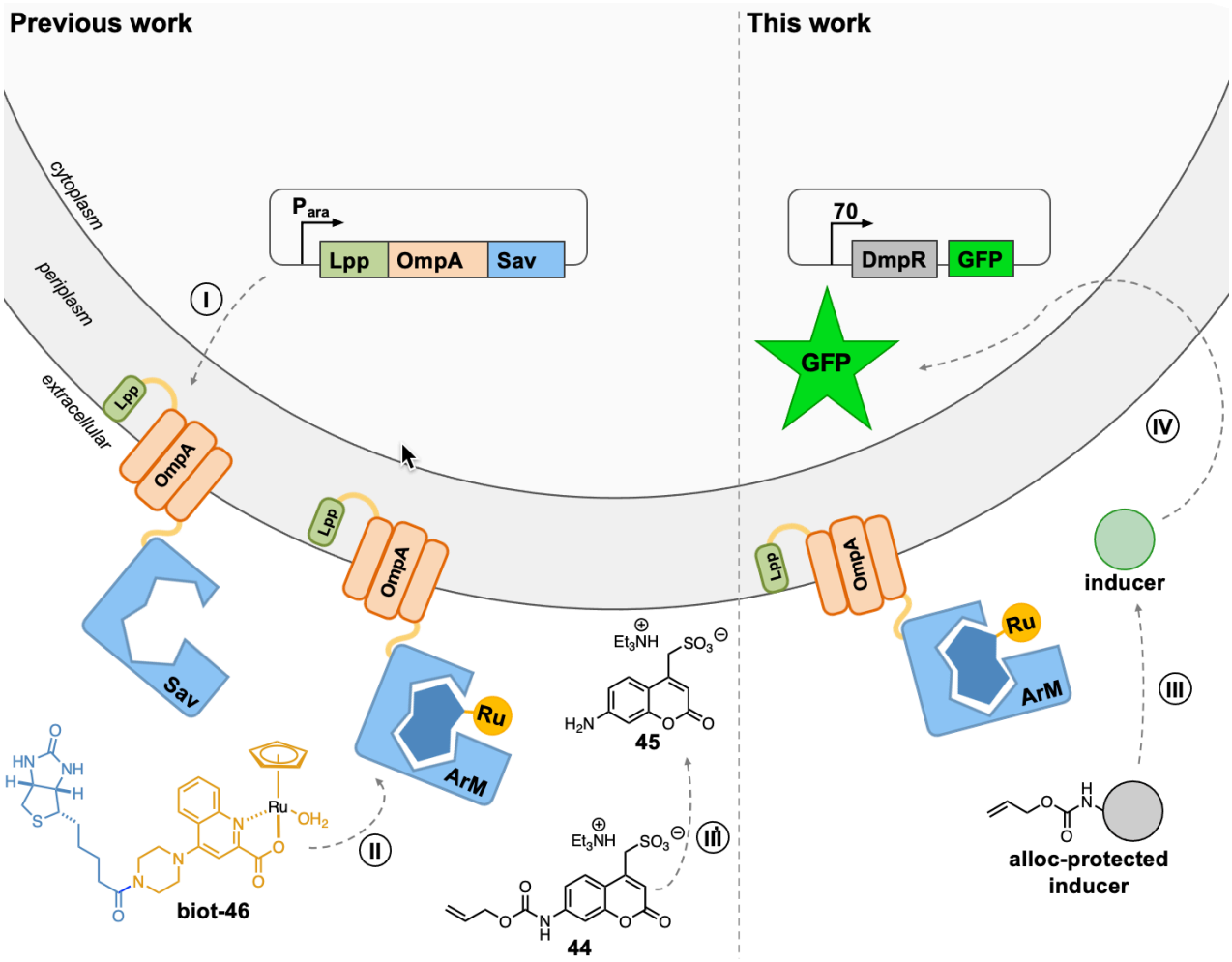


Figure 26. Comparison of the previous and the current strategies on optimizing ADAses using an *E. coli* surface-display strategy. I) expression, translocation and integration of the Lpp-OmpA-Sav construct on the outer membrane of *E. coli* cells. For clarity only one streptavidin (Sav) monomer is displayed. II) The biotinylated cofactor **46** binds to Sav to afford the ADAse. Previous work: III') uncaging of the protected substrate **44** to yield the fluorescent coumarin **45**. This work: III) uncaging of a protected inducer (grey circle) in the presence of the ADAse releases an inducer (green circle); IV) the inducer diffuses into the cytoplasm of *E. coli* cells and induces the over-expression of GFP.

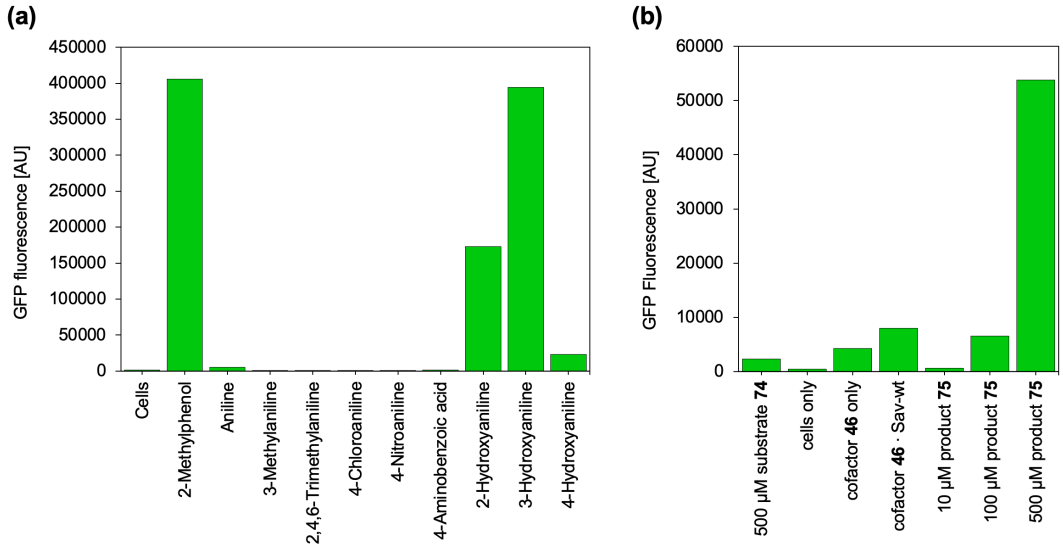


Figure 27. a) Induction capacity of various phenols and anilines for the DmpR regulator system. Reaction steps: i) *E. coli* DH5α cells containing the DmpR-GFP reporter plasmid were cultivated in LB-medium at 30 °C to a cell density of OD_{600} = 0.6. ii) Dilution of the cells to OD_{600} = 0.05-0.08, followed by addition of 500 µM caged inducer **74**. iii) Incubation at 30°C, 200 rpm shaking. iv) Analysis of the fluorescence intensity of the cells by flow cytometry (the median value of the fluorescence intensity is displayed). **b)** Performance of purified ADAses in the presence of reporter cells. Conditions: 500 µM substrate **74**, 5 µM ruthenium cofactor **46**, 2.5 µM Sav, 30 °C, 9 h.

Next, we combined our previously designed *E. coli* surface-displayed ADAses with the deprotection of caged inducer **74** and the subsequent expression of GFP, (Figure 28a). Control experiments revealed a modest background fluorescence for cells containing an empty vector (i.e. Lpp-OmpA plasmid without Sav gene), 2 µM cofactor **46** and 500 µM caged substrate 3, (Figure 28b). Gratifyingly, in the presence of the assembled ADAse = cofactor **46** · Sav-wt, a 1.7-fold increased fluorescence was observed. The observed fluorescence is comparable to a GFP expression level in the presence of ~50 µM inducer **75**. These findings clearly demonstrate the reliability of the surface-displayed ADAse for screening purposes.

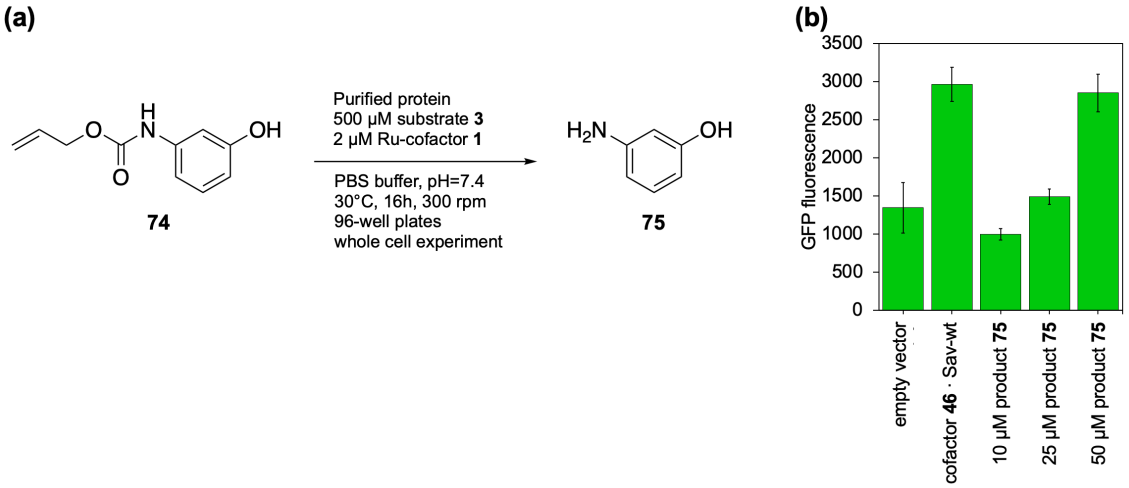


Figure 28. a) Reaction scheme and conditions for the allylic deallylation of the substrate **74** to the 3-hydroxyaniline **75**. Reaction conditions: 500 µM substrate **74**, incubation of cells with 2 µM cofactor **46**, 30 °C, 16 h. **b)** Control experiments to validate the screening strategy: catalysis using empty vector and Sav-wt was compared to cells expressing Sav-S112M-K121A spiked with 25 µM, 50 µM or 75 µM product **75**. GFP-fluorescence determined in a plate reader at: $\lambda_{ex.}$ = 475 nm, $\lambda_{em.}$ = 509 nm.

To validate the GFP-reporter strategy, we optimized the performance of the surface-displayed ADAse by simultaneously randomizing both S112 and K121 positions (Figure 29). Simultaneous saturation mutagenesis of two residues yields a library containing $20^2 = 400$ different mutants. The screening effort (i.e. the oversampling) can be reduced by a factor ~1.8 (1724 colonies instead of 3066 colonies with a sequence coverage of 95%) when using the degenerate codon DNK instead of NNK[19]. This library comprises of 17 different amino acids at each position (Gln, Pro and His are missing), leading to a total number of $17^2 = 289$ individual mutants.

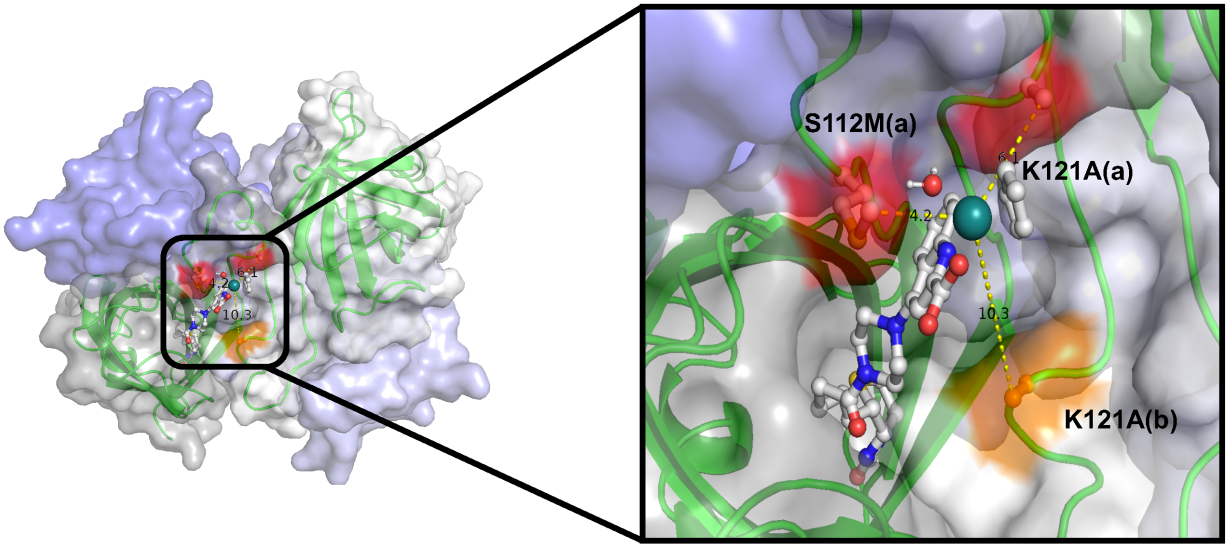


Figure 29. Crystal structure of the ADAse [CpRu(QA-Biot)(OH$_2$)] (46) · Sav S112M-K121A ([179] PDB 6FH8).The four monomeric units are displayed as transparent surface in grey, white, blue and light blue. The mutated positions (S112M^a and K121A^a) are highlighted in red and the position K121A^b of monomer B faces the active site is highlighted in orange (only in one of the two symmetry related monomers). The biotinylated ruthenium complex is displayed as ball and stick. Color code: C = grey, N = blue, O = red, S = yellow, H = white, Ru = dark green sphere. The cofactor **46** is anchored in the Sav monomer A (grey). The closest C_β residue to Ru (dark green sphere) is K121A (orange) of Sav monomer B (white).

The corresponding library was assembled by Gibson cloning of a Sav gene block library (90 base pair length, DNK codons in positions S112 and K121) into the Lpp-OmpA plasmid. Analysis of the library revealed a good statistical distribution of the different amino acids (Appendix B, Figure S7). However, the library contained a background of 42% (i.e. initially used template, frame shifts, stop codons, inserts, unclear sequencing results). Thus, a coverage of 93.8% is predicted for a total of 2762 screened individual colonies.[287] The 2762 colonies were tested and the 100 most active clones were sequenced. The 100 clones were retested in order to exclude false positives. Out of these 100 clones, the 10 best mutants (K121D, K121E, K121G, K121N, S112M-K121R, S112M-K121T, S112T-K121A, S112T-K121G, S112T-K121N and S112T-K121T) were selected for protein expression in the cytoplasm and *in vitro* characterization.

Since almost all double mutants were more active than Sav WT, we also tested the corresponding single mutants (i.e. S112M, S112T, K121A, K121G, K121N, K121R and K121T) in

order to identify potential synergetic effects. In addition, streptavidin wild-type (Sav WT), mutant S112M-K121A (i.e. the best mutant identified in previous work[179]) as well as the cofactor **46** alone were included in the *in vitro* catalysis validation (Figure 30). Comparison of the *in vivo* and *in vitro* ADAse activities revealed striking differences (Figure 30). Although variable expression levels may account for marked differences *in vivo* enzymatic activities,[179] past experience with Sav expression levels[288] led us to investigate an alternative hypothesis. Indeed, we surmised that fusing each Sav monomer (16.7 kDa) with a hydrophobic Lpp-OmpA insert (15.3 kDa) may affect the quaternary structure of surface-displayed Sav. Since two monomers of Sav make up the biotin-binding vestibule (Figure 29), the oligomeric nature of Sav may strongly affect which close-lying mutation of surface-displayed Sav (quaternary structure unknown) is indeed present in purified, tetrameric Sav samples.[274,289,290]

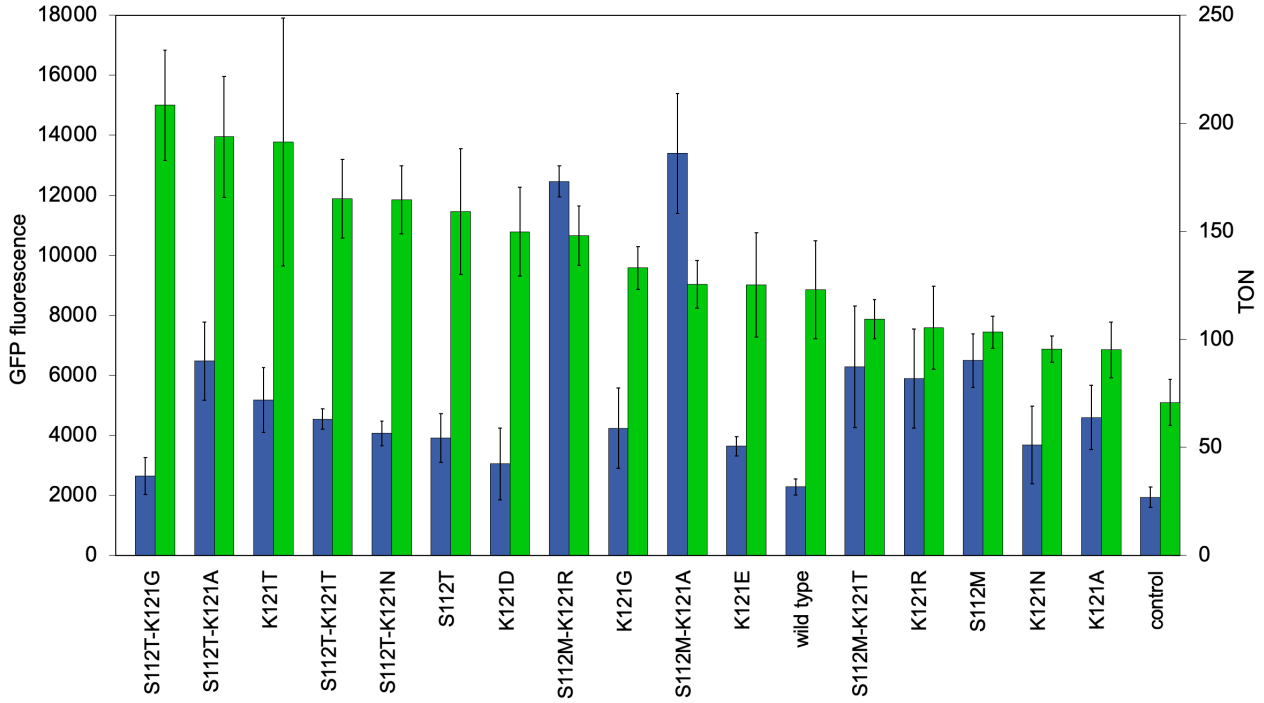


Figure 30. Directed evolution of artificial allylic deallylases for the uncaging of protected aniline 74. *In vitro* catalysis (blue bars): Substrate **74** (500 µM), cofactor **46** (1 µM), purified Sav isoform (2 µM biotin-binding sites), PBS-buffer (1x, pH 7.4), 30 °C, 18 h, 300 rpm shaking. Total turnover number (TON) determined by HPLC. Error bars = standard deviation from triplicate measurements. *In vivo* catalysis (green bars): 500 µM substrate **74**, incubation of cells with 2 µM ruthenium cofactor **46**, 30 °C, 16 h. GFP-fluorescence determined with a plate reader at: λ_{ex} = 475 nm, λ_{em} = 509 nm. Displayed values are normalized for the optical cell density (OD_{600}).

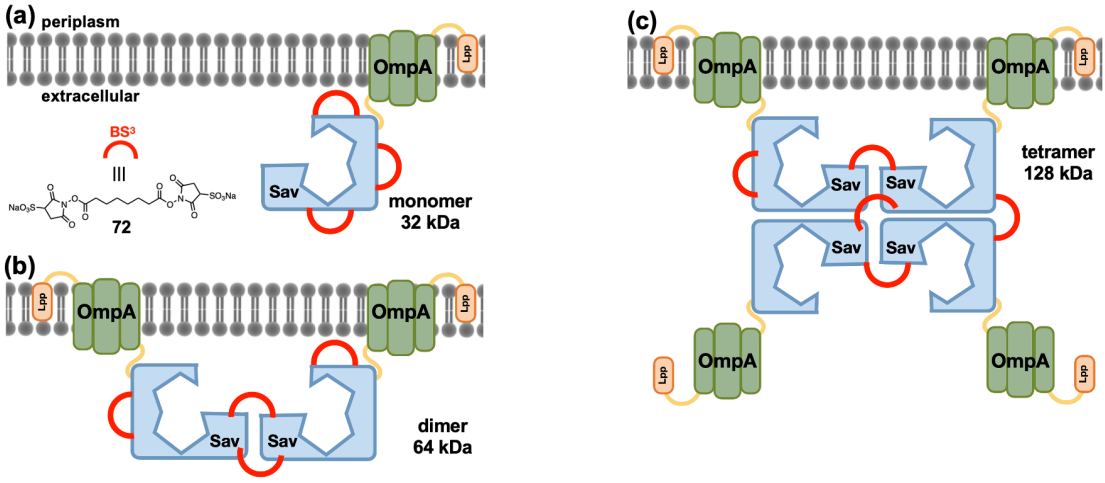


Figure 31. Schematic representation of Sav^{SD} using the Lpp-OmpA display system and the crosslinking approach.
a) Crosslinking with bis(sulfosuccinimidyl) suberate (**72**, BS^3), which mainly targets lysines, fixes the oligomeric state of the surface-displayed Sav^{SD}. Subsequent lysis of the cell membrane, releases the crosslinked constructs into solution that are analyzed by SDS-PAGE and western blot (see Figure 32). Depending on the surface-displayed quaternary structure of Sav^{SD}, three different sizes, **a)** 32 kDa, **b)** 64 kDa and **c)** 128 kDa can be expected for the monomer, dimer and tetramer. respectively.

To assess the quaternary structure of surface-displayed Lpp-OmpA-Sav, we set out to: i) treat the *E. coli* cells with a crosslinking agent, ii) lyse the cells and iii) analyze the protein-content by SDS-page and Western blot to identify the Sav-containing bands. Although surface display is a widely used strategy in protein expression and optimization, the quaternary structure of oligomeric, surface-displayed, proteins should to be addressed on a case-to-case basis, especially for oligomeric proteins.[286,291,292]

To minimize the adventitious assembly of homotetrameric Sav following cell lysis and workup, the cells were treated with bis(sulfosuccinimidyl) suberate (BS^3), a protein cross-linking agent,[245] to freeze the oligomeric state of Sav^{SD} prior to cell lysis (Figure 31). Exploratory crosslinking experiments highlighted the superiority of the anionic BS^3 *vs.* the commonly used glutaraldehyde.[245,246] The Sav^{SD} –as well as the empty vector as negative control– were expressed in *E. coli* at 25 °C for 24 h in shaking flasks. The cells were harvested and the pellets were washed with PBS before cross-linking with BS^3 (8 mM, 30 min, 25°C). The cross-linking was quenched by addition of 1 M Tris-buffer, the cell suspension centrifuged, the supernatant discarded and the cell pellet was resuspended in lysis buffer. The suspension was centrifuged and the supernatant was set aside for analysis. One part of the supernatant was loaded onto the gel (Figure 32, lane 6/9) and one part was denatured (SDS sample buffer, 95 °C, 30 min), (Figure 32, lane 7/10). The remaining cell pellet was resuspended in 8 M urea to ensure the recovery of any additional protein that might have been retained in the cell debris, centrifuged and the supernatant was subjected to analysis (Figure 32, lane 8/11). For comparison, purified WT Sav (pp-Sav) was subjected to cross-linking with BS^3. The resulting protein samples were analyzed by SDS-PAGE (Figure 32). To validate the BS^3

crosslinking strategy, control experiments using purified WT Sav were carried out. The pp-Sav samples that were first denatured and then crosslinked, mainly appear as monomers and partly as dimers (Figure 32, lane 5). In contrast, the denatured pp-Sav sample that was not crosslinked migrates as tetramer and higher oligomers on the SDS-PAGE (Figure 32, lane 2). Importantly, the pp-Sav samples that were crosslinked before the denaturation step retain the tetrameric structure (Figure 32, lane 4). We thus conclude that BS^3 crosslinking strategy permanently fixes the oligomeric state of Sav which is not affected by subsequent denaturation.

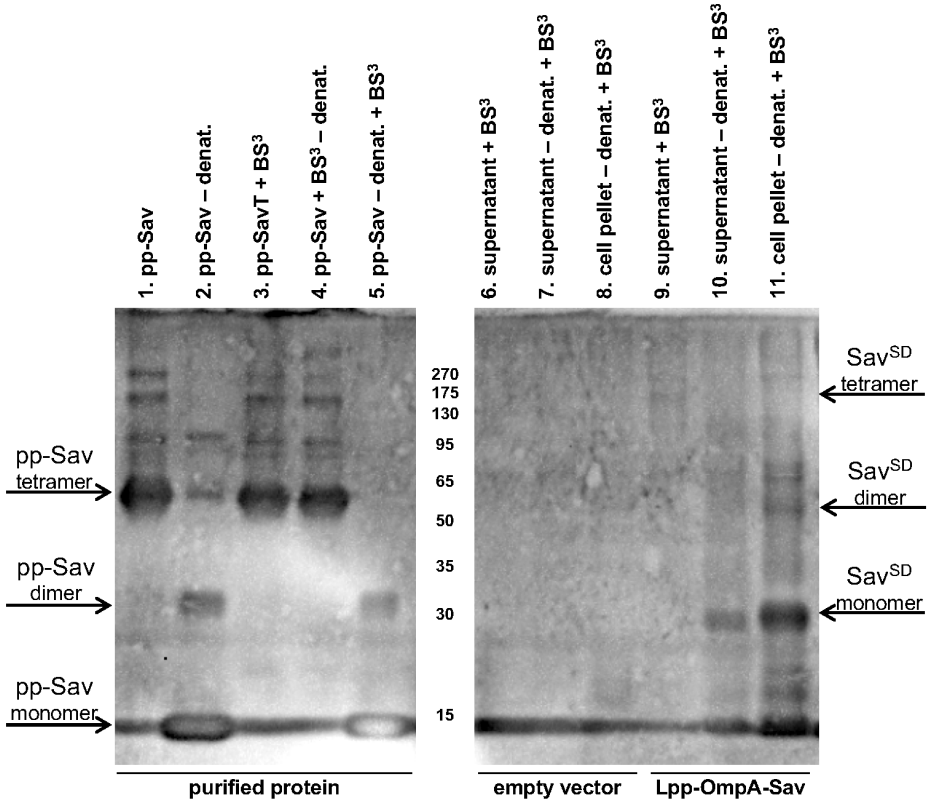


Figure 32. Western blot analysis of the quaternary structure of Sav^{SD}. Crosslinking of Sav^{SD}, followed by SDS-PAGE (14% polyacrylamide gel) and western blot (anti-Sav rabbit polyclonal antibody used at a 1:200 dilution). Except for lane 5, all samples were crosslinked with BS^3 prior to any additional treatment. All denatured samples were treated under the same conditions: SDS sample buffer, 95°C, 30 min. 1) pp-Sav, 2) denatured pp-Sav, 3) crosslinked pp-Sav, 4) crosslinked and then denatured pp-Sav, 5) denatured and then crosslinked pp-Sav, 6) empty vector, supernatant, 7) empty vector, denatured supernatant, 8) empty vector, denatured cell pellet 9) Sav^{SD}, supernatant, 10) Sav^{SD}, denatured supernatant, 11) Sav^{SD}, denatured cell pellet.

For Sav^{SD}, the electrophoresis results performed on a crosslinked Sav^{SD} sample, strongly suggest that Sav^{SD} is present as a monomer on the surface. Indeed, the denatured samples (Figure 32, lane 10/11) display a prominent band at about 32 kDa, which corresponds to the molecular weight of monomeric Sav^{SD} (Lpp-OmpA-Sav 31.925 kDA). In contrast, for the non-denatured Sav^{SD} sample (Figure 32, lane 9) faint bands at higher molecular weight are visible. Despite the chaotropic nature of the SDS-page analysis, the remarkable stability of homotetrameric Sav drives its oligomerization (tetramer and higher oligomers), as observed for the pp-Sav (Figure 32, lane 1-5).[256] Based on these qualitative observations, we hypothesize that the Lpp-OmpA-Sav is not

displayed as a homotetramer on the outer membrane of *E. coli*. Accordingly, care should be applied when optimizing the performance of surface displayed oligomeric enzymes, especially if their active site lies at the interface between monomers, as is the case for streptavidin (Figure 29).

3.3.4 Conclusion

An *E. coli* surface-displayed artificial allylic deallylase has been engineered to turn on a gene-switch leading to the upregulation of GFP. The ADAse activity was genetically optimized by simultaneous site saturation mutagenesis of positions S112 and K121. The most active surface-displayed ADAase (Sav- S112T-K121G) revealed a 1.7-fold improvement compared to the wild type Sav ADAse. Validation of surface-displayed ADAse activity, performed on purified Sav samples revealed marked differences between Sav^{SD} and pp-Sav data. Using pp-Sav, the best variant was Sav S112M-K121A with a 6-fold increased (and TON = 186) catalytic activity compared the WT Sav ADAse. The differences in activity between surface-displayed ArMs and purified samples– reflected by GFP fluorescence and 3-hydroxyaniline **75** detected by HPLC may be the result of multiple parameters: i) Sav^{SD} expression levels, the presence of side-products, including the formation of *N*-allylated 3-hydroxyaniline and iii) oligomeric nature of Sav^{SD}.

Acknowledgements

The authors are grateful to Dr. Johannes Rebelein and Dr. Alexandria Deliz Liang for helpful discussions, guidance and proofreading of the manuscript. Furthermore, Amanda Santos Kron, Nico Valerio Igareta and Juliane Klehr are acknowledged for the technical help. T.R.W. acknowledges the Swiss National Science Foundation (Grant 200020_182046), the NCCR Molecular Systems Engineering and the ERC (DrEAM - Advanced Grant 694424) for their generous financial support.

Author contribution

General concepts and idea by T.R.W. Cloning of the constructs and library creation by T.H., T.K and F.S., A.B. and F.S. performed the *in vivo* and *in vitro* experiments. A.B. performed the *in vitro* experiments and molecule characterizations via HPLC. J.V. designed and performed the crosslinking experiments. T.R.W., A.B. and J.V. wrote the manuscript. All authors reviewed the manuscript and the SI.

Supporting Information

The full supporting information can be found in Appendix B.

Conflict of Interest

The authors declare no conflict of interest.

3.4 REFERENCES

[156] U. T. Bornscheuer, G. W. Huisman, R. J. Kazlauskas, S. Lutz, J. C. Moore, K. Robins, *Nature* **2012**, *485*, 185–194.

[169] F. Schwizer, Y. Okamoto, T. Heinisch, Y. Gu, M. M. Pellizzoni, V. Lebrun, R. Reuter, V. Köhler, J. C. Lewis, T. R. Ward, *Chem. Rev.* **2018**, *118*, 142–231.

[171] M. E. Wilson, G. M. Whitesides, *J. Am. Chem. Soc.* **1978**, *100*, 306–307.

[179] T. Heinisch, F. Schwizer, B. Garabedian, E. Csibra, M. Jeschek, J. Vallapurackal, V. B. Pinheiro, P. Marlière, S. Panke, T. R. Ward, *Chem. Sci.* **2018**, *9*, 5383–5388.

[188] T. Vornholt, F. Christoffel, M. M. Pellizzoni, S. Panke, T. R. Ward, M. Jeschek, *Sci. Adv.* **2021**, *7*, 4208–4230.

[189] C. Stathopoulos, G. Georgiou, C. F. Earhart, *Appl. Microbiol. Biotechnol.* **1996**, *45*, 112–119.

[210] C. Streu, E. Meggers, *Angew. Chemie - Int. Ed.* **2006**, *45*, 5645–5648.

[211] T. Völker, F. Dempwolff, P. L. Graumann, E. Meggers, *Angew. Chem. Int. Ed. Engl.* **2014**, *53*, 10536–10540.

[212] H. Saburi, S. Tanaka, M. Kitamura, *Angew. Chemie - Int. Ed.* **2005**, *44*, 1730–1732.

[213] T. Völker, E. Meggers, *ChemBioChem* **2017**, *18*, 1083–1086.

[237] A. Charbit, J. C. Boulain, A. Ryter, M. Hofnung, *EMBO J.* **1986**, *5*, 3029–3037.

[238] R. Freudl, S. MacIntyre, M. Degen, U. Henning, *J. Mol. Biol.* **1986**, *188*, 491–494.

[239] J. A. Francisco, C. F. Earhart, G. Georgiou, *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 2713–2717.

[240] T. Chen, K. Wang, X. Chi, L. Zhou, J. Li, L. Liu, Q. Zheng, Y. Wang, H. Yu, Y. Gu, J. Zhang, S. Li, N. Xia, *Microb. Cell Fact.* **2019**, *18*, 70.

[241] S. Y. Lee, J. H. Choi, Z. Xu, *Trends Biotechnol.* **2003**, *21*, 45–52.

[242] J. Schüürmann, P. Quehl, G. Festel, J. Jose, *Appl. Microbiol. Biotechnol.* **2014**, *98*, 8031–8046.

[243] M. Park, J. Jose, S. Thömmes, J. Il Kim, M. J. Kang, J. C. Pyun, *Enzyme Microb. Technol.* **2011**, *48*, 307–311.

[244] T. Heinisch, F. Schwizer, B. Garabedian, E. Csibra, M. Jeschek, J. Vallapurackal, V. B. Pinheiro, P. Marlì, S. Panke, T. R. Ward, *Chem. Sci.* **2018**, *9*, 5383–5388.

[245] J. M. Shi, J. Pei, E. Q. Liu, L. Zhang, *PLoS One* **2017**, *12*, 1–13.

[246] R. Reinhart A.F., P. D. Bragg, *BBA - Biomembr.* **1977**, *466*, 245–256.

[247] M. Howarth, D. J. F. Chinnapen, K. Gerrow, P. C. Dorrestein, M. R. Grandy, N. L. Kelleher, A. El-Husseini, A. Y. Ting, *Nat. Methods* **2006**, *3*, 267–273.

[248] P. R. Johnson, M. Hochstrasser, *Trends Cell Biol.* **1997**, *7*, 408–413.

[249] L. E. Bird, *Methods* **2011**, *55*, 29–37.

[250] P. Bayer, A. Arndt, S. Metzger, R. Mahajan, F. Melchior, R. Jaenicke, J. Becker, *J. Mol. Biol.* **1998**, *280*, 275–286.

[251] J. Li, Q. Han, T. Zhang, J. Du, Q. Sun, Y. Pang, *Biotechnol. Reports* **2018**, *19*, e00261.

[252] X. F. Zhou, C. L. Zhang, X. P. Gao, W. L. Wang, Z. F. He, F. Y. Jiang, Y. L. Pang, J. H. Li, X. J. Ren, H. Bin Zhou, G. Q. Tan, J. X. Lyu, W. Wang, *AMB Express* **2020**, *10*, 1–9.

[253] R. Jürgen Dohmen, *Biochim. Biophys. Acta - Mol. Cell Res.* **2004**, *1695*, 113–131.

[254] H. Wang, Y. Xiao, L. Fu, H. Zhao, Y. Zhang, X. Wan, Y. Qin, Y. Huang, H. Gao, X. Li, *BMC Biotechnol.* **2010**, *10*, 14–24.

[255] M. P. Malakhov, M. R. Mattern, O. A. Malakhova, M. Drinker, S. D. Weeks, T. R. Butt, *J. Struct. Funct. Genomics* **2004**, *5*, 75–86.

[256] N. Humbert, A. Zocchi, T. R. Ward, *Electrophoresis* **2005**, *26*, 47–52.

[257] H. J. Davis, T. R. Ward, *ACS Cent. Sci.* **2019**, *5*, 1120–1136.

[258] M. T. Reetz, J. J. P. Peyralans, A. Maichele, Y. Fu, M. Maywald, *Chem. Commun.* **2006**, 4318–4320.

[259] M. Tomás-Gamasa, M. Martínez-Calvo, J. R. Couceiro, J. L. Mascarenãs, *Nat. Commun.* **2016**, *7*, 12538–12548.

[260] H. T. Hsu, B. M. Trantow, R. M. Waymouth, P. A. Wender, *Bioconjug. Chem.* **2016**, *27*, 376–382.

[261] T. Völker, E. Meggers, *ChemBioChem* **2017**, *18*, 1083–1086.

[262] H. M. Key, P. Dydio, D. S. Clark, J. F. Hartwig, *Nature* **2016**, *534*, 534–537.

[263] W. J. Song, F. A. Tezcan, **2021**, *346*, 1525–1528.

[264] A. R. Grimm, D. F. Sauer, T. Polen, L. Zhu, T. Hayashi, J. Okuda, U. Schwaneberg, *ACS Catal.* **2018**, *8*, 2611–2614.

[265] R. Blomberg, H. Kries, D. M. Pinkas, P. R. E. Mittl, M. G. Grütter, H. K. Privett, S. L. Mayo, D. Hilvert, *Nature* **2013**, *503*, 418–421.

[266] P. Srivastava, H. Yang, K. Ellis-Guardiola, J. C. Lewis, *Nat. Commun.* **2015**, *6*, 7789–7797.

[267] A. Ilie, M. T. Reetz, *Isr. J. Chem.* **2015**, *55*, 51–60.

[268] O. F. Brandenberg, R. Fasan, F. H. Arnold, *Curr. Opin. Biotechnol.* **2017**, *47*, 102–111.

[269] S. Kille, C. G. Acevedo-Rocha, L. P. Parra, Z. G. Zhang, D. J. Opperman, M. T. Reetz, J. P. Acevedo, *ACS Synth. Biol.* **2013**, *2*, 83–92.

[270] T. R. Ward, *Acc. Chem. Res.* **2011**, *44*, 47–57.

[271] M. Jeschek, R. Reuter, T. Heinisch, C. Trindler, J. Klehr, S. Panke, T. R. Ward, *Nature* **2016**, *537*, 661–665.

[272] U. Markel, D. F. Sauer, J. Schiffels, J. Okuda, U. Schwaneberg, *Angew. Chemie* **2019**, *131*, 4500–4511.

[273] R. Obexer, A. Godina, X. Garrabou, P. R. E. Mittl, D. Baker, A. D. Griffiths, D. Hilvert, *Nat. Chem.* **2017**, *9*, 50–56.

[274] M. Park, J. Jose, S. Thömmes, J. Il Kim, M. J. Kang, J. C. Pyun, *Enzyme Microb. Technol.* **2011**, *48*, 307–311.

[275] M. T. Reetz, J. D. Carballeira, *Nat. Protoc.* **2007**, *2*, 891–903.

[276] M. Hestericová, T. Heinisch, L. Alonso-Cotchico, J.-D. Maréchal, P. Vidossich, T. R. Ward, *Angew. Chemie* **2018**, *130*, 1881–1886.

[277] J. Jose, R. Bernhardt, F. Hannemann, *ChemBioChem* **2001**, *2*, 695–701.

[278] J. Jose, S. Von Schwichow, *ChemBioChem* **2004**, *5*, 491–499.

[279] T. Peschke, K. S. Rabe, C. M. Niemeyer, *Angew. Chemie* **2017**, *129*, 2215–2219.

[280] R. Rubini, C. Mayer, *ACS Chem. Biol.* **2020**, *15*, 3093–3098.

[281] Y. Okamoto, R. Kojima, F. Schwizer, E. Bartolami, T. Heinisch, S. Matile, M. Fussenegger, T. R. Ward, *Nat. Commun.* **2018**, *9*, DOI 10.1038/s41467-018-04440-0.

[282] V. Shingler, T. Moore, *J. Bacteriol.* **1994**, 1555–1560.

[283] V. Shingler, M. Bartilson, T. Moore, *J. Bacteriol.* **1993**, 1596–1604.

[284] S. L. Choi, E. Rha, S. J. Lee, H. Kim, K. Kwon, Y. S. Jeong, Y. H. Rhee, J. J. Song, H. S. Kim, S. G. Lee, *ACS Synth. Biol.* **2014**, *3*, 163–171.

[285] K. K. Kwon, D. H. Lee, S. J. Kim, S. L. Choi, E. Rha, S. J. Yeom, B. Subhadra, J. Lee, K. J. Jeong, S. G. Lee, *Sci. Rep.* **2018**, *8*, 2659–2668.

[286] C. F. Earhart, *Methods Enzym.* **2000**, *326*, 506–516.

[287] M. T. Reetz, J. Sanchis, *ChemBioChem* **2008**, *9*, 2260–2267.

[288] T. Vornholt, F. Christoffel, M. M. Pellizzoni, S. Panke, T. R. Ward, M. Jeschek, *Sci. Adv* **2021**, *7*, 4208–4230.

[289] J. A. Francisco, C. F. Earhart, G. Georgiou, *Biochemistry* **1992**, *89*, 2713–2717.

[290] J. Jose, T. F. Meyer, *Microbiol. Mol. Biol. Rev.* **2007**, *71*, 600–619.

[291] C. Stathopoulos, G. Georgiou, C. F. Earhart, *Appl. Microbiol. Biotechnol.* **1996**, *45*, 112–119.

[292] E. van Bloois, R. T. Winter, H. Kolmar, M. W. Fraaije, *Trends Biotechnol.* **2011**, *29*, 79–86.

4

ULTRAHIGH-THROUGHPUT DIRECTED EVOLUTION OF ARMS

4.1 INTRODUCTION

As previously mentioned, directed evolution of biological systems, such as enzymes, requires extensive screening of vast sequence spaces. Traditionally directed evolution is done in an iterative manner over multiple generations and has succeeded in creating new functions or achieving higher activities. But despite directed evolution nowadays being a common tool to optimize a given enzymatic system, the understanding of how and why a certain mutation influences the scaffold and thus leads to an improved activity is still unclear. Because of the dynamic nature of proteins, even remote mutations can have pronounced effects on the catalytic performance of enzymes.[4,293,294] A remarkable example is the directed evolution of aspartate aminotransferase that showed 10^6-fold increase in catalytic activity, but most of the mutated residues were non-active-site amino acids.[295]

Epistasis (Greek for "breaking", "stopping") is historically a term used to describe the masking of a specific phenotype linked to one gene by the presence of another gene. It is a genetic phenomenon where multiple mutations are co-dependent on each other: the effect of a given mutation is highly influenced by the genetic background of where it appears. In general, three types of epistasis are distinguished:[296]

(i) Additive epistasis: in the case of a double mutation, the effect of the double mutation is the sum of the single mutations.

(ii) Positive epistasis: the new double mutation's effect is more than the sum.

(iii) Sign epistasis: the double mutation can have a negative influence on the overall activity.

Therefore, ideally, when performing directed evolution, multiple positions should be targeted at the same time to allow identifying positive or sign epistasis. However, the screening and analysis throughput available to most researchers to date is limited. Thus, oftentimes one still relies on iterative genetic optimization, which is a time consuming, tedious and often times incomplete process. A seminal paper by John Maynard Smith as early as 1970, illustrated this problem by describing the term of the protein fitness landscape. He illustrated the protein landscape as every possible sequence of a protein containing *n* amino acids, where each neighboring protein sequence only differs by exactly one amino acid. In a 'Gedankenexperiment' he demonstrated the magnitude of the possibilities: if a small protein of 100 amino acids is considered, there are 20^{100} sequences possible, a number larger than the number of atoms present in the universe. Nature probed this space for millions of years and selected a tiny fraction of that number with improved biologically-relevant activities.[297]

The protein fitness landscape therefore strongly influences the evolution of the system. The easiest landscape is the so called single-peaked 'Fujiyama' landscape (Figure 33a) where each iteration will be a step in the uphill path leading to one single peak. However, the reality is much more complicated and rather resembles a so called multi-peaked 'Badland' (Figure 33b). Local optima are like traps that can only be overcome if a decrease in activity is acceptable during at least one round of evolution –or if multiple positions can be screened simultaneously.[298]

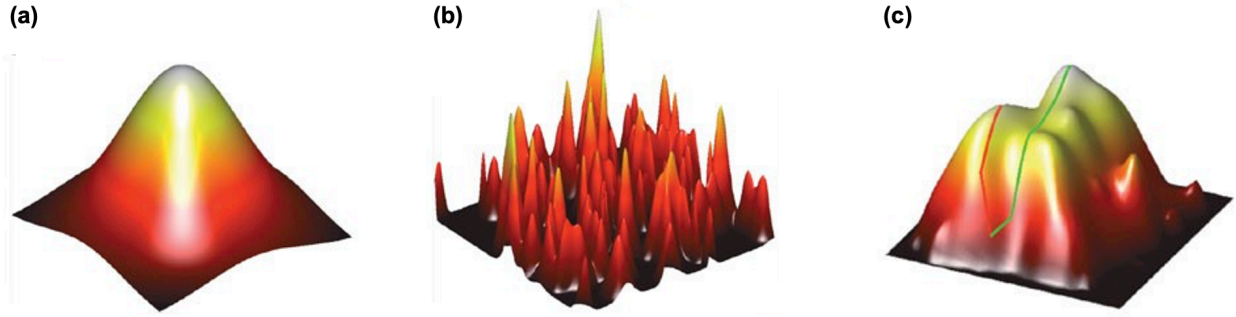


Figure 33. Protein fitness landscapes. A proteins fitness is the measure of how the genetic change influences the target function. **a)** Ideal 'Fujiyama' landscape where each added mutation is beneficial and will be a step further to the optimum. **b)** Rugged 'Badland' landscape where multiple optima are possible and traversing from a local maximum to another higher local maximum requires temporary loss in function. **c)** Comparison of two iterative pathways which lead to different local maxima depending on the pathway taken.[293]

Figure 33c depicts how iterative screening can identify two different local maxima. Consequently, screening the whole sequence space should lead to the ultimate maximum; however, the available screening and analysis throughput so far is limited, making true exploration of whole

such fitness landscapes impossible. With a back-of-the-envelope calculation, if the screening time for one position requires 1 week, the effort increases to 4 weeks, 128 weeks and >80 years for two, three and four positions screened simultaneously. Here, an ultrahigh-throughput method for the rapid multi-site target directed evolution of ArMs is outlined, allowing a significant decrease of this screening time to one week on average to be able to screen a bigger sequence space in one round (Figure 34).

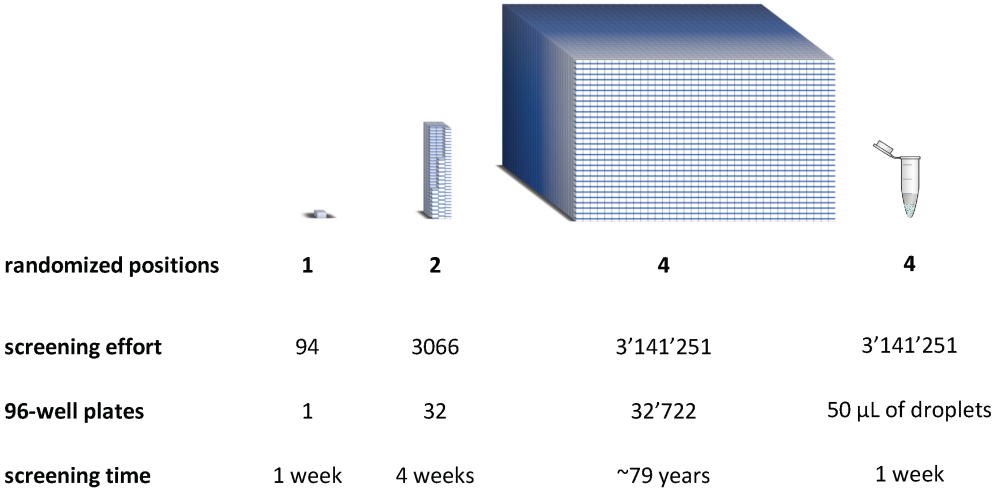


Figure 34. (Repetition of Figure 1) **Schematic representation of the screening effort per mutated position using an NNK library.** An NNK library at one position has 32 possible codons encoding for the twenty amino acids. This corresponds to a screening effort of 94 colonies to achieve a theoretical library coverage of 95%. This effort increases exponentially if two or more positions are screened simultaneously. Screening four positions would require about 80 years, considering that eight 96-well plates are screened per week. In comparison, screening the same library in double emulsions using microfluidic tools would require about one week of work.[19]

The following sections will give an overview on initial attempts to develop a medium-throughput screening assay based on agar plates (Chapter 4.2), followed by the method development of an ultrahigh-throughput screening assay based on double-emulsion microfluidics (Chapter 4.3), and its first application to a quadruple mutant library (Chapter 4.5).

4.2 Initial Approach: Agar Plate Assay

The screening throughput is most often lost in the compartmentalization step. For MTP assays compartmentalization consists in the localization of the genotype (DNA, *E. coli*) in the same well as its corresponding phenotype (catalytic activity). One major bottleneck lies in the step where the library (colonies on an agar plate) needs to be transferred to a 96-well plate and the number of transferred colonies limits the screening throughput. Moreover, the consumption of reaction media and valuable reaction components including cofactor and substrate are high: about 0.5 L reaction medium for 500 colonies in 96-well plates compared to ~1 mL for an agar plate assay (Figure 35).

We investigated a straightforward screening assay for ArMs based on agar plates where growth and expression of the library, as well as the reaction monitoring can be performed on a single agar plate. Turner and coworkers used such a screening assay for the detection of colored by-products on agar-plates to measure the *ω*-Transaminase activity.[17] Furthermore, they developed a label-free screening platform that allowed the monitoring of bacterial colonies on agar plates by desorption electrospray ionization coupled with ion mobility mass spectrometry imaging (DiBT-IMMS) and demonstrated the potential of this method on two different enzymes, ammonium lyase and P450 monooxygenase.[299]

Inspired by these types of screening, we envisioned similar monitoring of the catalytic reaction by fluorescence imaging. The deprotection reaction of aminocoumarin **44** (Figure 35b) and the aminophenol **74** (Figure 35c) mentioned in chapter 2 and 3 respectively, served as a starting point for the method development.

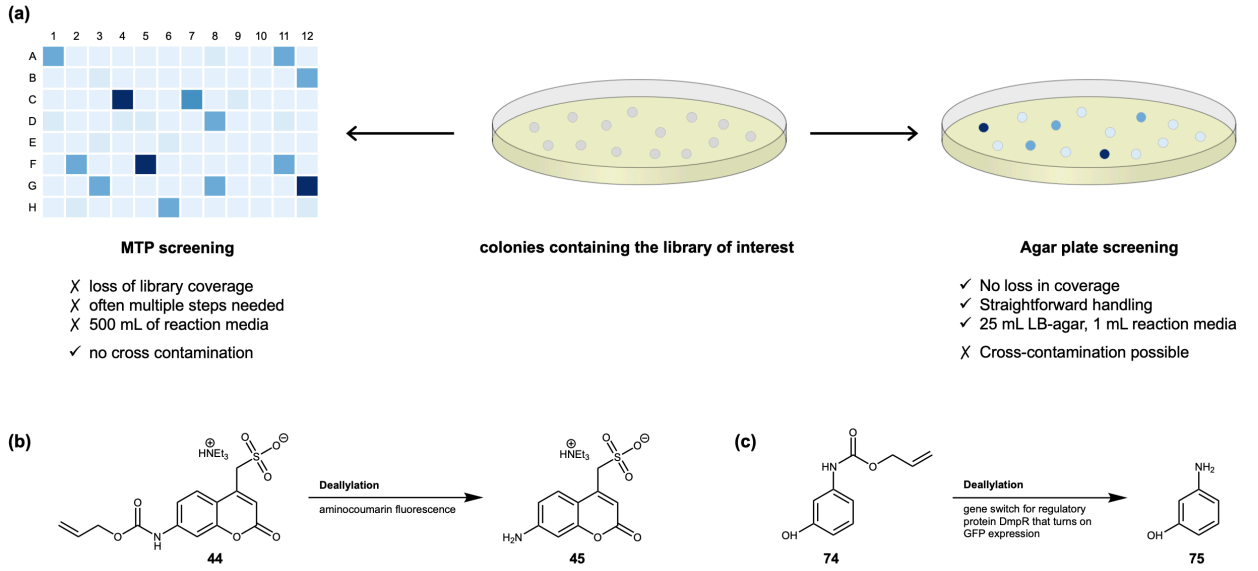


Figure 35. Agar plate assay. a) Schematic representation and advantages of an agar plate assay. **b)** Deprotection of fluorescent aminocoumarin **44**. **c)** Deprotection of aminophenol **74** which turns on the overexpression of GFP.

Initially, the conditions of *E. coli* growth and Sav expression on agar plates were optimized (see Appendix C, Figure S21-23). Briefly, Top10 (DE3) cells were plated on an LB-agar plate containing the inducer for Sav expression. After expression, cofactor **46** (2-20 µM, 500 µL) was sprayed onto the agar plate and incubated at room temperature for 1-2 h to allow cofactor uptake. Subsequently, the substrate (200 µM, 250 µL) was sprayed onto the agar plate too and incubated for the reaction period. The reaction was monitored at regular intervals by fluorescence imaging and the images were analyzed using Image J. To determine if differences in reactivity between different mutants are detectable using this assay, a 1:1 ratio of Sav-wt and the good Sav-MA mutant were plated and analyzed.

For the deprotection of aminocoumarin **44**, the coumarin fluorescence was measured. A clear difference in activity between cells expressing Sav and empty vector was observed. However, the difference in catalytic activity observed by fluorescence imaging was too small to enable the distinction between Sav-MA and Sav-wt. (see Appendix C, Figure S24 for detailed reactivity)

The deallylation of aminophenol **75** that turns on GFP expression, yielded better results. At high catalyst loading, clear differences in activity could be detected by fluorescence imaging. But analysis of ~30 colonies that showed the highest fluorescence signal by Sanger sequencing gave only modest results. Only a slightly higher percentage of Sav-MA was observed in the brightest colonies (Figure 36). We hypothesize that the issue might be the effect of cross-contamination which, over time, leads to an overall same fluorescence intensity.

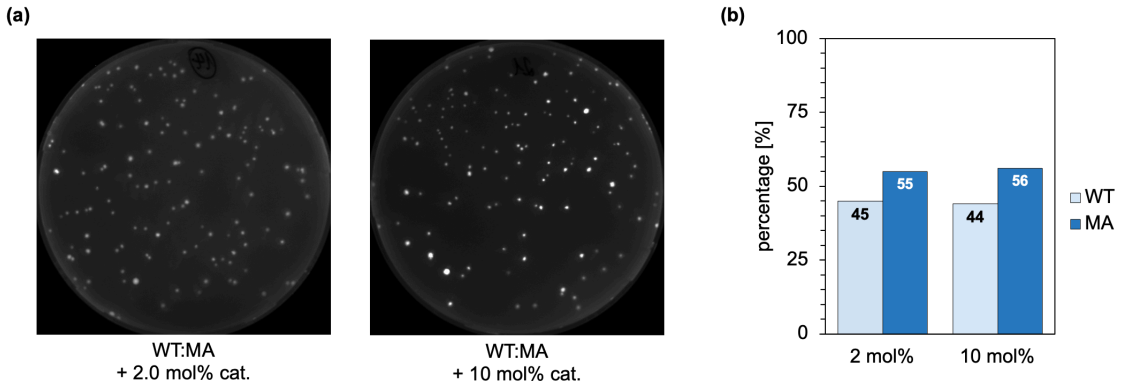


Figure 36. Agar plate screening of Sav-wt ***vs.* Sav-MA in a 1:1 ratio for the deprotection of aminophenol 74**. **a)** fluorescence imaging of two conditions that were used for the analysis with Image J. **b)** Sanger sequencing analysis of ~30 colonies with brightest fluorescence intensities.

Reverting back to compartmentalization with distinctly separate entities, we opted for miniaturization. Microfluidics-based encapsulation of *E. coli* cells in double emulsions (DEs) and subsequent fluorescence activated cell sorting (FACS), allowed the fast compartmentalization and analysis of live *E. coli* cells. The method development (Chapter 4.3) and first applications of the ultrahigh-throughput screening assay (Chapter 4.5) are outlined below.

4.3 DE Screening – Method Development

4.3.1 Reaction Selection

The heart of any screening assay is a reaction that enables efficient monitoring of the enzymes' reactivity. For the *in vivo* directed evolution in double emulsion droplets, the first step consisted of identifying a suitable reaction combining three key characteristics:

(i) *Biocompatibility*: the catalytic reaction needs to be oxygen tolerant, feasible in an aqueous medium, compatible with cellular components (e.g. high concentrations of thiol) and must not interact with any other molecule in the cell. This is to ensure that directed evolution can be carried out using live *E. coli* cells without harming the cofactor or the *E. coli*.

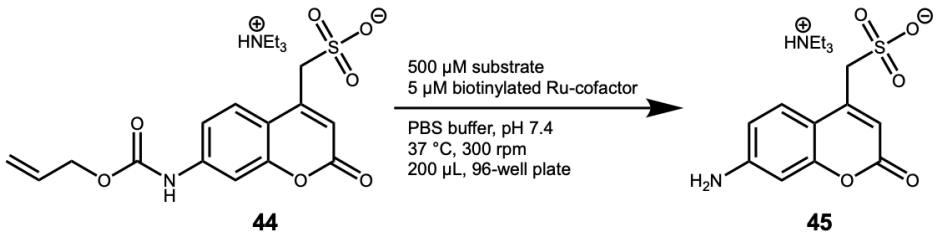
(ii) *Molecular retention*: molecules need to be retained inside droplets to avoid cross-over between individual droplets. To create a tight genotype-phenotype linkage, the reaction components need to remain localized inside the water droplet of the double emulsion. Compounds which are too lipophilic will accumulate in the oil shell and therefore hinder catalysis and FACS may not be feasible. The partition coefficient (log *D*) is a measure for a compound's solubility between two different solvents (water and octanol). A log D value lower than -2.5 predicts a good retention inside water droplets.[300,301]

(iii) *Fluorescent product*: the product needs to be fluorescent to enable the analysis of the reaction by FACS.

Bioorthogonal reactions such as metathesis and hydroxylation using ArMs previously reported within the Ward-group were initially envisaged. However, the synthesis of a profluorescent substrate turning into a fluorescent product was not trivial. Moreover, the poor retention of either the substrate or the product was a major challenge (Appendix C, Figure S26).

The allylic deprotection of the allyl-carbamate protected coumarin **44** using the artificial deallylase fulfilled all of these conditions (Scheme 15). The reaction is biocompatible, and both the substrate **44** as well as the fluorescent product **45** are water soluble and are well retained inside the droplets. Building on previous experience of *in vivo* directed evolution of the ADAse using surface-displayed Sav, and directed evolution campaigns using periplasmic expression of Sav, the ADAse was selected as the reaction of choice.

Table 6 summarizes the *in vitro* catalytic activities of key Sav-variants used throughout this thesis. Wild type monomeric Sav (Sav-wt) and two previously identified promising mutants S112M-K121R (Sav-MR), and S112Y-K121R (Sav-YR), were compared. Moreover, with the goal of screening single chain dimeric Sav (scdSav), scdSav-wt, the dimeric version of a quadruple mutant (scdSav-SARK) and the monovalent isoform of the same mutant (scdSav-mv2-SARK) were tested.[190] To confirm that the reaction is not affected by oxygen, the reaction was also carried out under inert atmosphere in degassed PBS buffer (Table 6, entries 8-14). It can be concluded that oxygen has no detrimental effect on the catalysis in the long-term. However, it should be noted, that the initial rates are higher under air-free conditions, but plateau at longer reaction times.



Scheme 15. Deallylation of the caged allyl carbamate-aminocoumarin **44** to the aminocoumarin **45** by an ADAse.

Table 6. Catalytic activity of different purified ADAses at various time points for the catalytic uncaging of substrate **44** to product **45**.

Entry	Complex/ArM	Catalyst loading [mol%]	TON				
			0.5 h	1.0 h	2.5 h	8.0 h	16 h
1[a]	substrate + cofactor **46**	1	0.3	3.2	65	68	76
2[a]	Sav-wt	1	0.1	0.7	19	27	46
3[a]	Sav-MR	1	0.3	2.2	57	64	63
4[a]	Sav-YR	1	0.2	1.3	51	62	58
5[a]	scdSav-wt	1	0.1	1.0	14	18	25
6[a]	scdSav-SARK	1	0.1	0.9	25	35	53
7[a]	scdSav-mv2-SARK	1	0.2	3.0	43	49	62
8[b]	substrate + cofactor	1	0.5	12	78	80	80
9[b]	Sav-wt	1	0.3	9.0	56	60	60
10[b]	Sav-MR	1	0.8	29	53	58	59
11[b]	Sav-YR	1	0.7	30	56	60	58
12[b]	scdSav-wt	1	0.2	8.6	49	56	60
13[b]	scdSav-SARK	1	0.5	14	54	58	57
14[b]	scdSav-mv2-SARK	1	0.5	13	63	65	65

Reaction conditions: substrate **44** (500 µM), cofactor **46** (5 µM), Sav (2.5 µM), PBS (pH 7.4), 0.5% DMF, 37 °C, 300 rpm. Conversion determined by the measurement of the product **45** fluorescence intensity.

Further, the same reaction conditions were applied to cells displaying Sav on the surface (Figure 37, grey), or expressing Sav in the periplasm (Figure 37, blue), and bearing an additional mNectarine plasmid (Figure 37, orange). It is evident, that the periplasmic expression yields higher catalytic activities. Cells expressing mNectarine additionally performed on average 1.5-fold worse. However, the differences between the different variants were still apparent, which made us confident, that the reaction is suitable for the screening in double emulsion droplets. Based on these initial activities, a microfluidics-based method for the screening of ADAse activity in double emulsion droplets with downstream FACS analysis was developed.

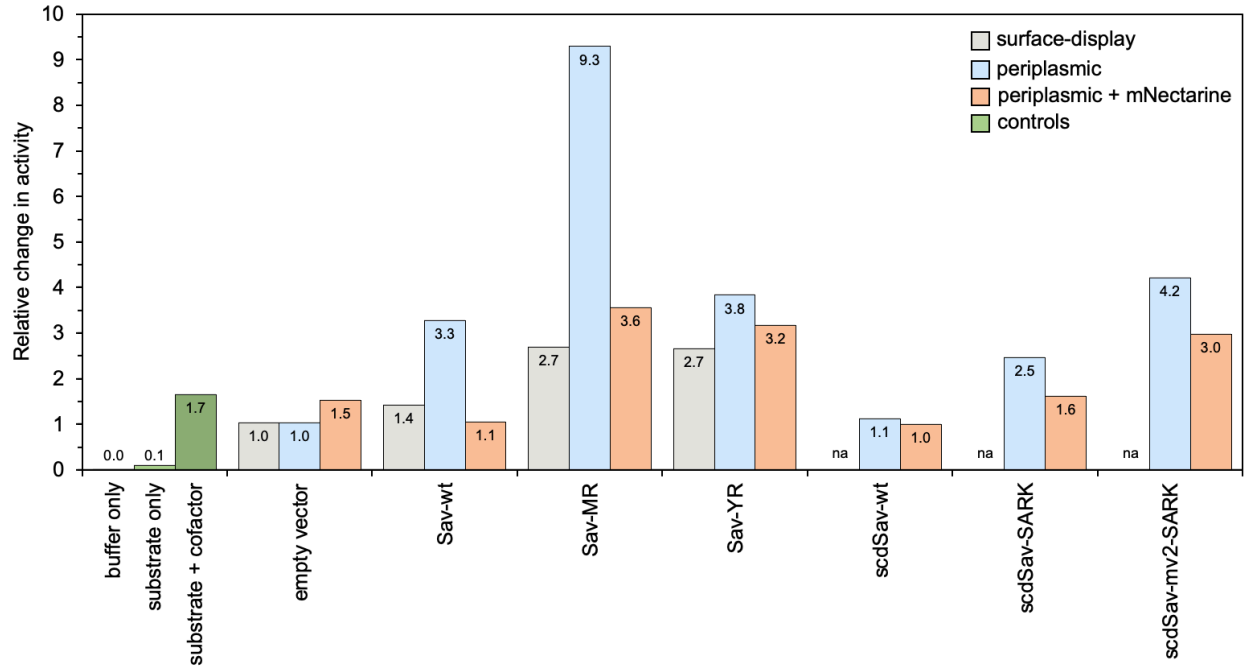


Figure 37. Catalytic activity of different ADAses screened using *E. coli* cells. Surface-displayed Sav (grey), periplasmic expression (blue) and the addition of the mNectarine plasmid (orange) were compared.

4.3.2 Double Emulsion Production and Cell Encapsulation

This work was mainly performed by Ariane Stucki or in collaboration with Ariane Stucki.

To encapsulate the reaction components, a PDMS chip bonded to a PDMS-coated glass slide was used. It contained two inlets for the inner aqueous phase (IA), one inlet for the oil phase (OP), one inlet for the outer aqueous phase (OA) and one outlet (Figure 38). The chip involved two flow focusing junctions to enable the sequential formation of water-in-oil droplets followed by the water-in-oil-in-water droplet. Flow rates between 0.5 and 2 µL min^{-1} were used for the inner aqueous phases and the oil phase and flow rates up to 5 µL min^{-1} were used for the outer aqueous phase. To ensure double-emulsion formation on one single chip, the OA and the outlet channels were coated with 2.5% polyvinyl alcohol (PVA).[302] Using this chip, monodisperse DEs with a diameters comprised between 15-18 µm, corresponding to an internal volume between 1.5 and 3.0 pL, were produced at a throughput of 4000-6000 Hz (Figure 38). The oil phase consisted of a fluorinated oil (HFE 7500 with 2% 008-FluoroSurfactant). PBS was used as the inner aqueous phase (IA) and PBS with 0.5% SDS was used as the outer aqueous phase (OA). The two inlets ensured, that *E. coli* incubated with the cofactor **46**, and the substrate **44** could be added via different inlets and only got in contact at the time of encapsulation. The components of the inner aqueous phases were prepared in double the final concentration, as a 1:1 ratio was achieved at the droplet production junction.

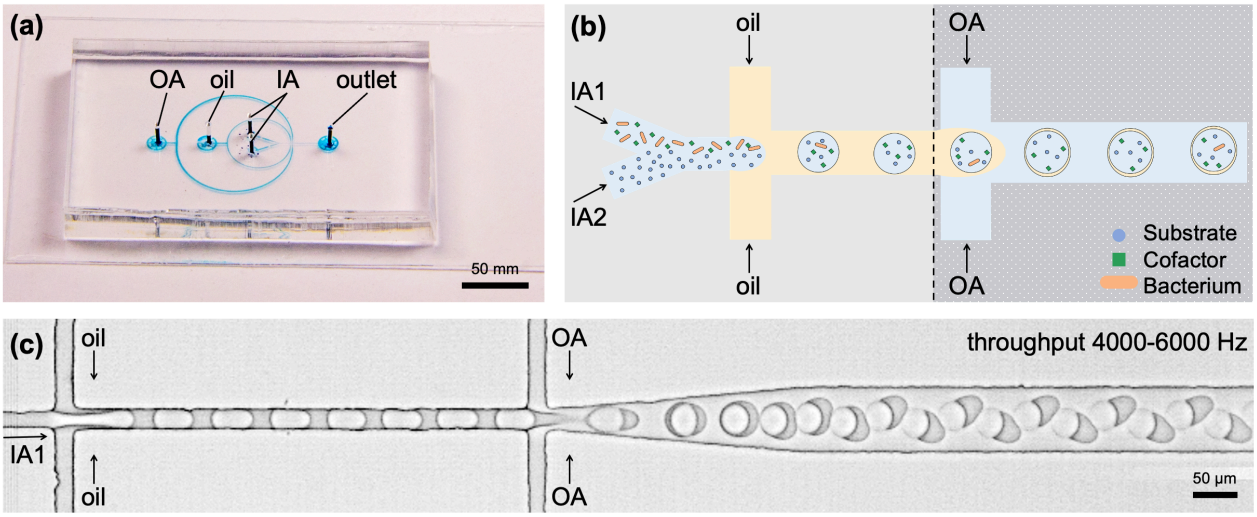


Figure 38. Design of the microfluidic chip for the production of DEs. a) PDMS chip bonded to a glass slide that is used for the DE production. Channels are colored using a blue dye solution for better visibility. Scale bar: 50 mm. **b)** Schematic representation of the chip design with four inlets for cells and cofactor (IA1), substrate (IA2), oil and the outer aqueous phase. The hydrophilic coated area is depicted in dark grey. **c)** Micrograph of DE production. Water-in-oil droplets are produced at the first flow focusing junction. A second emulsification step leads to the formation of water-in-oil-in-water double emulsions. The inner aqueous phase channels have a diameter of 10 µm, the oil channels have a diameter of 15 µm, and the outer aqueous phase channels have a diameter of 20 µm. Scale bar: 50 µm.

With big libraries in mind, where droplet production over an extended period of time is essential, the production stability was investigated (Figure 39a). In the first minutes, the produced droplets vary considerably in size. However, the production stabilizes significantly afterwards and remains stable ($^{+}/_{-}$ 3% limit) for the time period of investigation. Once produced, the droplets retain their shape, and stay intact and monodisperse for at least 24 h (Figure 39b). The reported numbers and production stabilities were obtained with the initial chip design that produced droplets with ~50 µM diameter. The chip design was subsequently optimized to afford smaller reaction volumes reported here and to improve the signal-to-noise ratio and FACS compatibility. The reported stability and monodispersity are analogous for the smaller droplets produced using the new chip design.

To ensure that only one single *E. coli* cell is encapsulated in one droplet, the cell solution was used at a dilution of OD_{600} = 0.1. While the assumption that a solution with an optical density of OD_{600} = 1 contains ~10^9 cells/mL, an OD_{600} = 0.1 corresponds to a final encapsulation of ~0.2 cells/droplet (Figure 40a). This has the effect, that 76% of droplets produced are “empty droplets” (i.e. contain no cells), 22% contain one *E. coli*, and the probability that multiple *E. coli* are encapsulated within one droplet is lower than 3%. Considering the high production rate, the chip design allows the encapsulation of ~1000-1300 *E. coli* per second, making the compartmentalization of large libraries feasible in a short time.

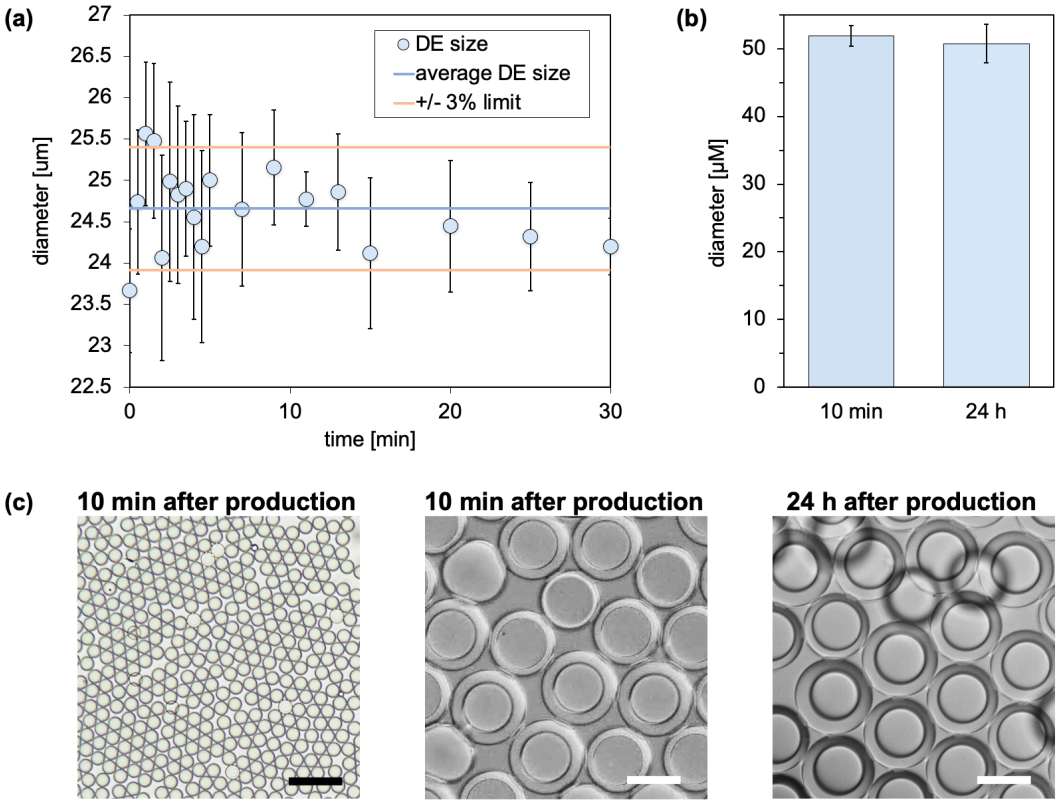


Figure 39. Droplet stability and monodispersity. a) Droplet production stability over the initial 30 min of production. **b)** Average size of droplets measured 10 min and 24 h after production. Average values and standard deviation were calculated by measuring 50 droplets. **c)** Micrographs of the droplet population 10 min after production and 24 h after production. Black scale bar: 200 µm, white scale bar = 50 µm.

To confirm the Poisson distribution, cells expressing GFP (GFP was introduced to make them easily visible for counting) were encapsulated in droplets. The droplets were trapped on a second PDMS chip (Figure 40b) with 614 hydrodynamic traps, and counted manually.[303] Counting droplets containing a viable cell (moving GFP signal) between 10-30 min after production, the expected distribution was observed. This not only confirmed the statistical distribution of the *E. coli* (~22%), but also confirmed that the cells were viable during this period. However, incubation overnight diminished the number of droplets with a GFP signal and only about 8% cells/droplet could be observed. It is assumed that the majority of cells die because of the nutrient-free PBS-buffer they are in. This assumption is further backed by the poor regrowing efficiency of encapsulated *E. coli* (Table 7). Additionally, the experiment was repeated with cells expressing Sav. After expression the cells were labeled with a fluorescent biotinylated Atto-dye (Figure 40c). This confirmed the presence of Sav, as well as the cell viability in the first hours after the encapsulation (Appendix C, Figure S30).

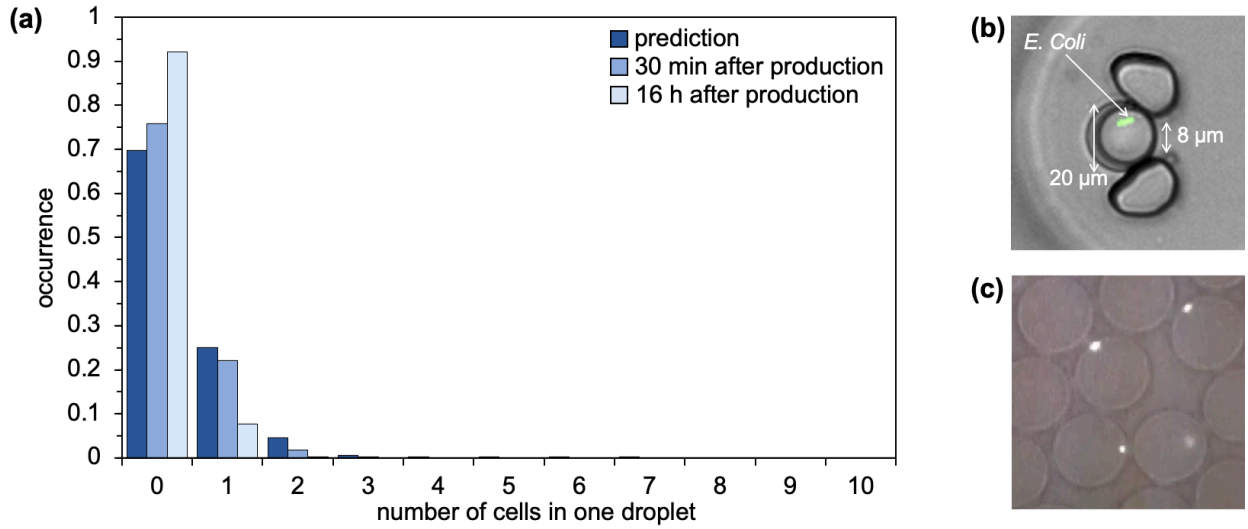


Figure 40. Encapsulation of single *E. coli* cells. a) Poisson distribution of encapsulated *E. coli*. **b)** DE containing a GFP-labeled *E. coli* trapped on a trapping chip. **c)** DE containing *E. coli* expressing Sav labeled with the biotinylated fluorescent dye Atto-565.

4.3.3 Flow cytometry analysis & FACS

In a next step the ADAse was encapsulated and monitored via flow cytometry and fluorescence microscopy. Initial experiments were conducted using double emulsions with either (i) purified protein combined with cofactor and substrate, or (ii) *E. coli* cells expressing variants with known activities together combined with cofactor and substrate.

First, the coumarin fluorescence intensity and stability was assessed by encapsulating different concentrations of substrate **44** and product **45**. Different DE populations with only substrate **44** or product **45** were produced. The flow cytometry analysis of the different substrate **44** and product **45** populations az different time points confirmed that both substrate and product can be encapsulated in DEs and are well retained within the droplets (Figure 53).

The DEs were produced on chip, collected in an Eppendorf tube and incubated off-chip at 37 °C, without shaking. To analyze the different populations, the DEs (0.5 µL) were resuspended in the OA (300 µL), manually mixed and analyzed on a flow cytometer. The presence of mNectarine fluorescence (λ_{ex} = 560 nm, λ_{em} = 580 nm) enabled the discrimination of DEs with cells from the ones without cells. This population was further analyzed for coumarin fluorescence (λ_{ex} = 405 nm, λ_{em} = 450 nm).

4.3.4 Purified Protein in Double Emulsions

In a next step, purified protein was encapsulated together with cofactor **46** and substrate **44**. Two different DE samples, one containing Sav-wt and one Sav-YR, were produced. After production, the samples were mixed in a 1:1 ratio and analyzed by flow cytometry at regular intervals. The two populations have clearly distinguishable catalytic activities (Figure 41a).

Moreover, trapping droplets with encapsulated Sav-YR, and analyzing them *via* fluorescence microscopy over the whole period, revealed that the reaction reaches maximum fluorescence intensity within the first three hours (Figure 41b/c). With these findings, and considering the cell viability, it was estimated that the sorting should be conducted between 3-4 h after the encapsulation.

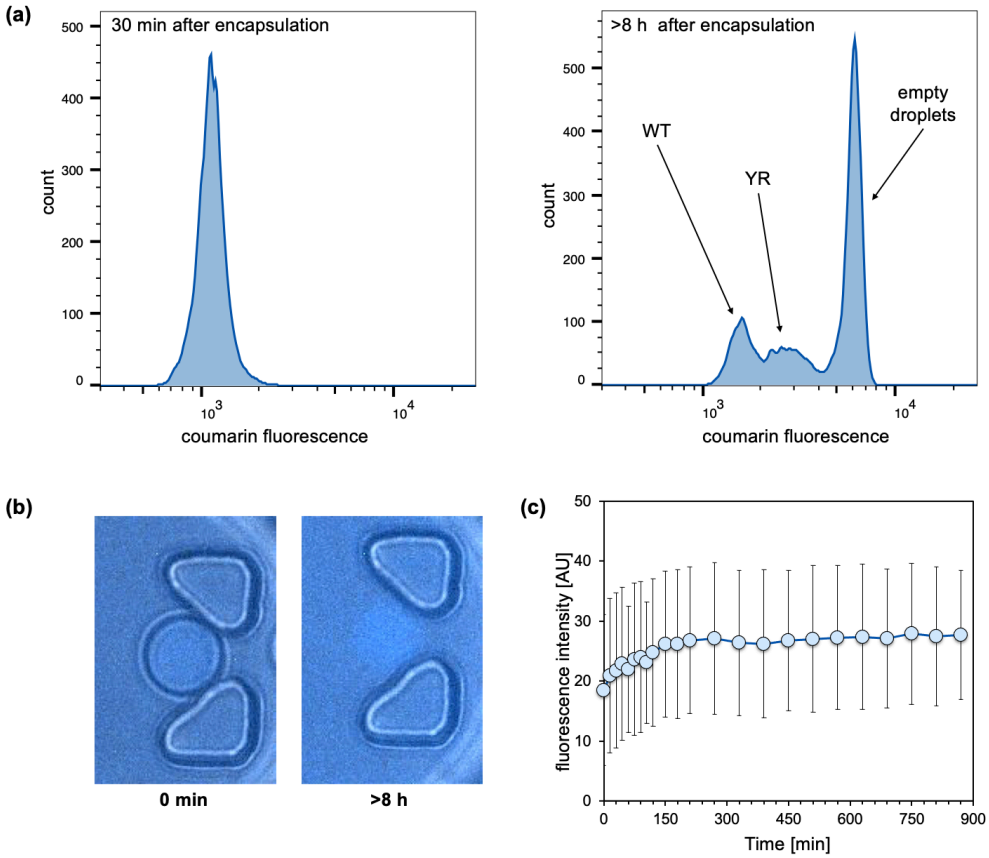


Figure 41. Flow-cytometry and fluorescence microscopy observation of the catalytic activity of encapsulated purified protein. a) Separately encapsulated Sav-wt and Sav-YR were mixed in a 1:1 ratio after encapsulation and observed over 8 h. **b)** Sav-YR encapsulated in DEs and observed *via* fluorescence microscopy over 8 h. **c)** Summarized time point measurement obtained *via* fluorescence microscopy.

4.3.5 *E. COLI* IN DOUBLE EMULSIONS

In a next step the effect of *E. coli* on the substrate's integrity was investigated. DEs co-encapsulating *E. coli* expressing Sav-wt together with substrate **44** (500 µM) but without the addition of the cofactor, did not change the fluorescence intensity over time (Figure 42a). Comparing the fluorescence intensity of empty DEs and DEs containing a cell over 10 h, revealed that the substrate is not converted into the product only by the presence of a cell. The fluorescence intensity stayed constantly low over the observed period and no substrate leakage was observed.

In the next control, *E. coli* not expressing Sav were incubated with the cofactor **46** and encapsulated with substrate **44** (Figure 42b). Here, both empty DEs and DEs containing a cell, behaved very similarly and an increase in coumarin fluorescence intensity was observed over time

in both cases. Inspection of the fluorescence intensity of the empty DE population, an evident background fluorescence is observed, which is even more pronounced when *E. coli* cells expressing Sav are used. For the screening assay, this posed a challenge, as it revealed that free cofactor present in the droplet also contributes to catalysis. This could lead to two adverse effects: (i) false positive results, and (ii) the background reaction being so high that individual ADAse activities cannot be distinguished.

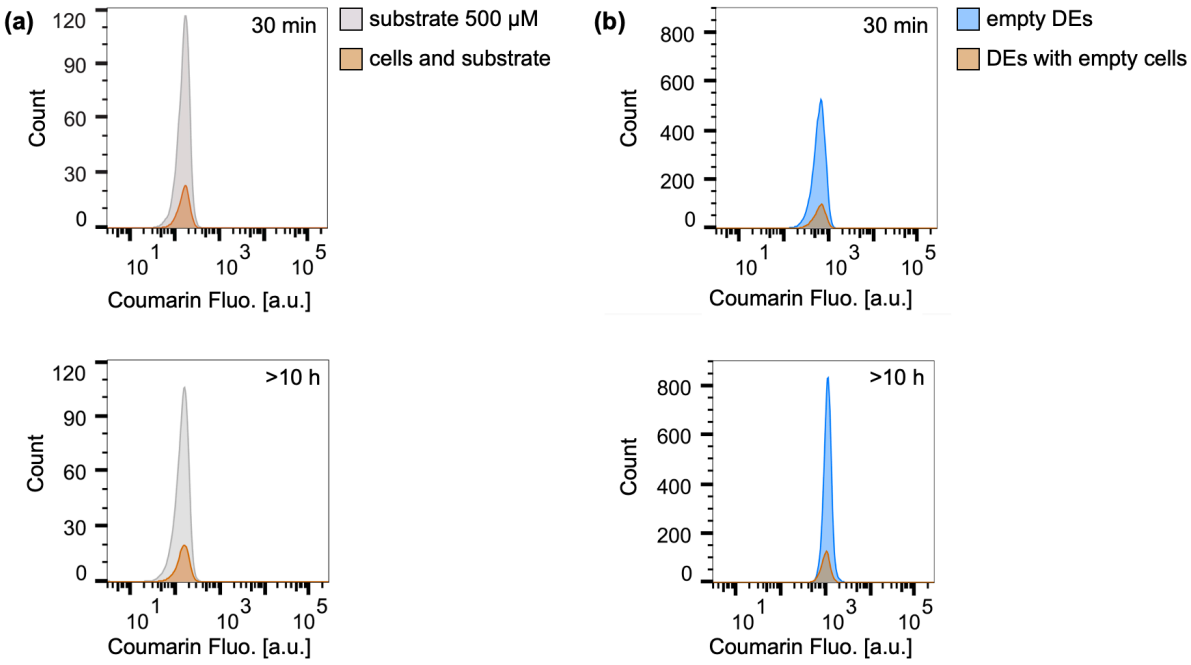


Figure 42. Substrate and product stability in the presence of *E. coli* cells. a) Coumarin fluorescence intensity distribution of DEs containing 500 µM substrate and Sav-wt cells, obtained 30 minutes and 18 hours after encapsulation. **b)** Coumarin fluorescence intensity distribution of DEs encapsulating substrate, cofactor and cells without plasmid for Sav expression, obtained 30 minutes and 10 hours after encapsulation.

To analyze this effect further, a 1:1 mixture of GFP-labeled Sav-wt and mNectarine-labeled Sav-MR were incubated with cofactor **46**. The sample was split into two aliquots:

(i) Sample 1 was promptly encapsulated in DEs together with the substrate **44**.

(ii) Sample 2 was incubated with the cofactor **46** for 30 min on ice and washed three times with PBS to remove excess cofactor prior to encapsulation in DEs with substrate **44**.

The coumarin fluorescence intensity of both DE populations was analyzed using a flow cytometer over 6.5 h (Figure 43). The washing step indeed significantly reduced the background activity (Figure 43). However, the whole coumarin fluorescence intensity profile shifted to lower intensities, with the effect that the difference between different mutants was very small. Moreover, the intensities were so low that it proved to be more difficult to sort since the machine used for FACS was less sensitive than the flow cytometer used to analyze the DE.

For comparison, Figure 43c displays a close-up view of the populations with GFP and mNectarine signals. The ratio distribution of Sav-wt *vs.* Sav-MR, show a corresponding 24-fold enrichment for Sav-MR in the top 5% of the DEs. This clearly indicates that the difference in catalytic activity is measurable even in the presence of the background signal caused by the activity of the free cofactor. Moreover, since a fluorescent protein is used as an internal readout to reveal the presence *E. coli*, the empty DEs can be easily sorted out. With this semi-quantitative control, the feasibility of the procedure is warranted.

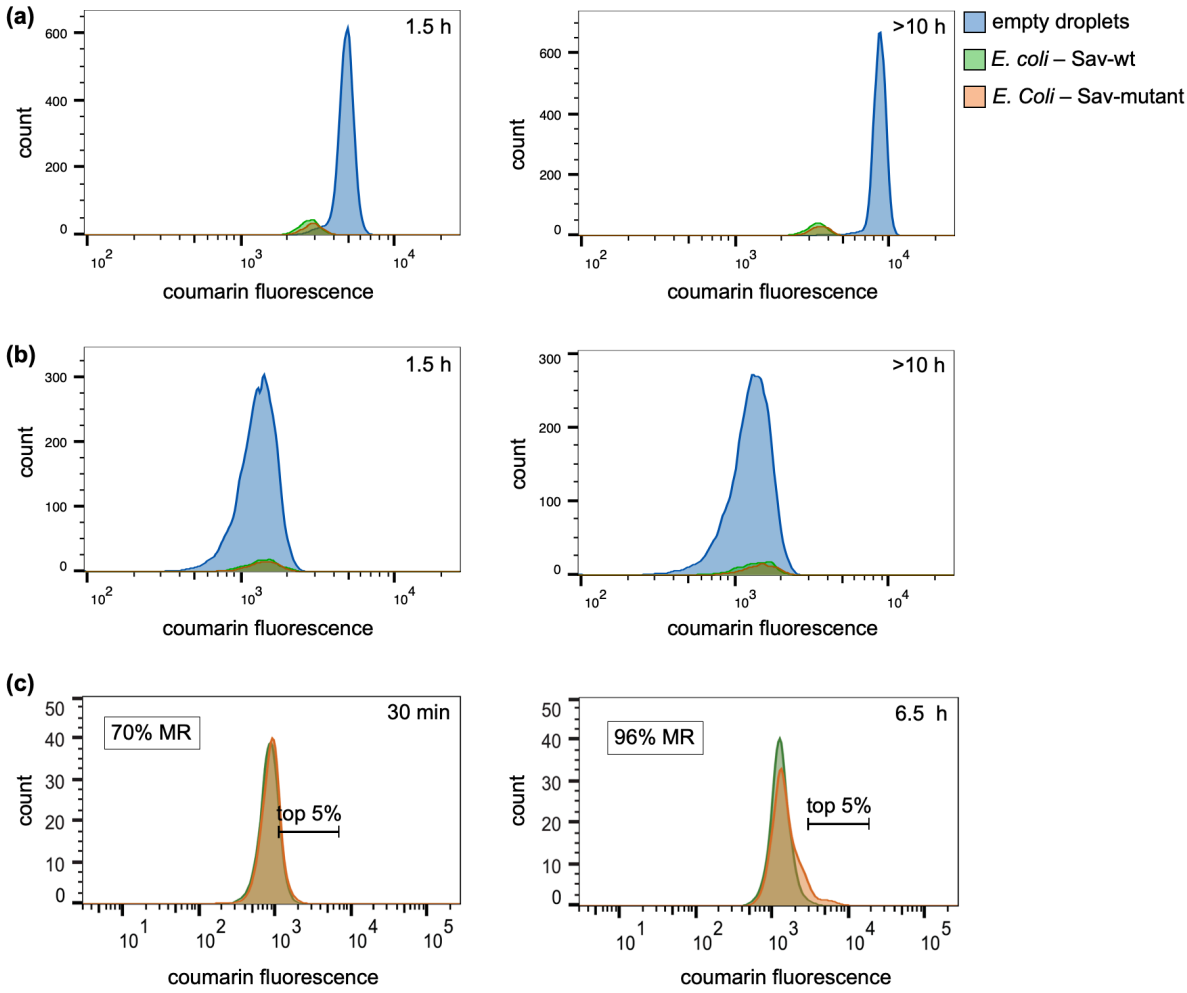


Figure 43. Reducing the background. Coumarin fluorescence intensity distribution of DEs containing GFP-labeled Sav-wt (green), mNectarine-labeled Sav-mutant (orange) and the empty DEs (blue) at two different time points. **a)** Direct encapsulation of the cells incubated with cofactor produces a prominent background peak. **b)** Encapsulation of cells incubated with cofactor and excess cofactor removed by washing. **c)** Sav-wt *vs.* Sav-MR distribution over a period of 6.5 h.

4.3.6 Retrieving the Genetic Information

Directed evolution can only take place if the genotype-phenotype linkage is maintained, and the genetic information of the hits with high coumarin fluorescence intensity can be retrieved. Since the catalysis is performed *in vivo*, regrowing the sorted populations on an agar plate after FACS

would be ideal. However, as briefly presented above, the cell viability in the selected reaction conditions is a major challenge. We compared cells not overexpressing any protein (empty vector), cells expressing Sav and cells expressing human carbonic anhydrase (hCA) for their regrowing efficiency.

To do so, first different *E. coli* cells were encapsulated in DEs. After the encapsulation, the DEs were incubated at 37 °C without shaking and a sample was taken at different time points (30 min, 4 h, 8 h, overnight). To estimate the regrowing efficiency, different techniques were applied: (i) the DEs were directly plated without any further treatment, (ii) the DEs were broken using 1*H*, 1*H*, 2*H*, 2*H*-perfluorooctanol (a reagent used to disrupt the water-oil-water interface and thus break the droplets), resuspended in fresh LB-medium, incubated for 30 min at 37 °C and then plated, or (iii) the plasmid was extracted and transformed into electrocompetent Top10 (DE3) *E. coli* cells.

It could be seen that *E. coli* cells which expressed Sav formed significantly lower amounts of colonies than the empty vector or the hCA controls (Table 7). As it is reported that cells encapsulated in nutritional medium tend to regrow better, we tested minimal medium instead of PBS as the inner aqueous phase. Rich media such as LB, SOC etc. were not suitable options because of their autofluorescence. Using a minimal medium, the regrowth efficiency could be slightly improved. However, the number of colonies that regrow is so low, that the coverage needed for analyzing bigger libraries, is jeopardized.

One approach that yielded slightly more colonies, was the transformation of the extracted plasmid into electrocompetent *E. coli* cells. However, also this approach did not prove efficient enough to obtain statistically-relevant numbers of colonies.

DNA extraction of the sorted sample and PCR amplification of the extracted plasmid was investigated in a next attempt. Initial experiments using the parent plasmid DNA revealed that a minimum concentration of 0.8 pg/µL was necessary for efficient PCR amplification. However, the sorted samples were not easily amplified because of the relatively low number of cells sorted. In a standard sorting experiment roughly 20’000-30’000 events were sorted for the highest gate, corresponding to 20’000-30’000 single *E. coli*. Assuming that 1-10 plasmids are present in one *E. coli*, roughly 3.5-35 ag of plasmid DNA is present in one cell. Thus 20’000 sorted events correspond to ~0.07 pg plasmid DNA in a sorted sample, which is well below the limit of 0.8 pg/µL determined by PCR amplification of the parent plasmid.

Table 7. Comparison of different methods for the retrieval of genetic information after the sort.

Entry	Sample	CFU / 5 µL DEs	
		30 min	20 h
1[a]	Sav – no encapsulation	>1000	>1000
2[b]	empty vector	>1000	>500
3[b]	hCA	>1000	>1000
4[b]	Sav	~100	<5
5[c]	empty vector	>1000	>1000
6[c]	Sav	~25	~5
7[d]	empty vector	>1000	>1000
8[d]	Sav	~50	<5
9[d]	Sav – unsorted	na	>100
10[d]	Sav – sorted	na	20
11[e]	empty vector	na	>1000
12[e]	Sav – unsorted	na	>150

Conditions: a) no encapsulation, 500 µL of an OD_{600} = 10-6. b) no additional treatment, directly plated after encapsulation. c) the droplets were broken by the addition of 1*H*, 1*H*, 2*H*, 2*H*-perfluorooctanol, and resuspended in LB medium at 37 °C and 30 min before plating. d) The cells were resuspended in PBS + 1% glucose before encapsulation in DEs. e) Transformation of plasmid extracted from DEs. f) The cells were resuspended in minimal medium before encapsulation in DEs.

Optimization of the PCR amplification process consisted in the investigation of different plasmid extraction kits (Macherey-Nagel, Sigma-Aldrich, Monarch kit by NEB, LiOAc precipitation), and different polymerases (Q5, Pfu, KAPA HiFi, Taq). Plasmid extraction using the Monarch kit by NEB, combined with Taq polymerase yielded most efficient DNA amplification (see Appendix C, Figure S33 for details). With this protocol, enough DNA was obtained to perform either (i) TOPO cloning, followed by transformation and Sanger sequencing of single colonies, (ii) Next Generation Sequencing using the Illumina platform (NGS), or (iii) Nanopore sequencing.

4.3.7 Analysis Options

Analysis of the PCR amplified material needs an efficient sequencing technique that can distinguish single base changes in the sequence. Moreover, the technique needs to offer similarly high throughputs as achieved through encapsulation and sorting to cover the same number of variants screened.

4.3.7.1 TOPO cloning and Sanger Sequencing

In an initial effort, TOPO cloning in combination with Sanger sequencing was envisioned. TOPO cloning utilizes the enzyme DNA topoisomerase I, which naturally recognizes the specific sequence 5'-(C/T)CCTT-3' on both ends of double-stranded DNA, digests the DNA at this position, unwinds the DNA and religates it again to the thymidine base at the 3' end. For TOPO cloning, a linearized TOPO vector, which has the topoisomerase covalently attached to its 3' ends, is used. Mixing this TOPO vector with a PCR product that has a 3'-adenine overhang (Taq or Pfu

polymerases) enables the ligation of both fragments by the already present topoisomerase I (Figure 44).

We used this approach to amplify the Sav gene using Taq polymerase, followed by TOPO cloning to achieve a circular plasmid. Transformation of the TOPO cloning product resulted in colonies that were picked and sequenced by Sanger sequencing. However, the cloning and transformation efficiency was relatively low (~10-50 colonies for a sorted sample), thus the results obtained through this method were not statistically useful, as a good coverage of the library size was not reached.

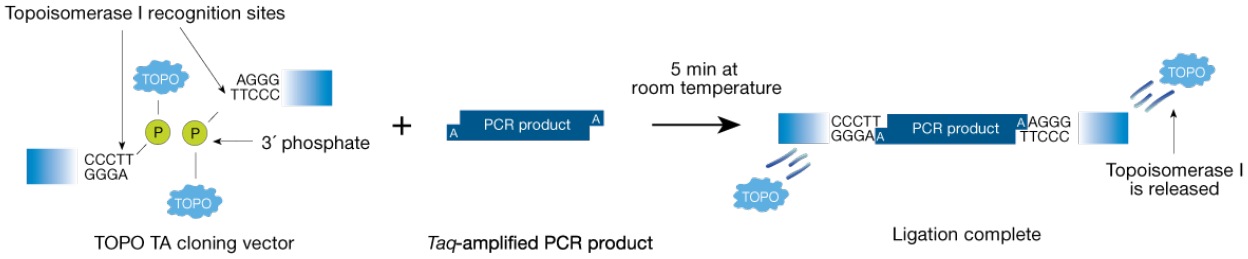


Figure 44. Schematic representation of TOPO cloning using a Taq-amplified PCR product.[304]

4.3.7.2 Next Generation Sequencing (NGS) – Illumina Sequencing

Illumina sequencing is the most widespread next generation sequencing (NGS) sequencing technology.[305] It is based on the strand termination technology used for Sanger sequencing but utilizes a glass surface chip preloaded with forward and reverse primers. Addition of small DNA fragments with a complementary sequence to the primers on-chip,[306] leads to so called bridge PCR amplification of the fragment DNA in a first step.[307] Small local colonies of identical DNA strands are produced on-chip (Figure 45a) so that amplicons from one DNA fragment will only cluster in one single physical spot on the array. In a next step a separate set of sequencing primers is added. Sequencing is then initiated by the addition of the four dNTPs each bearing a different fluorophore and a polymerase. The modified dNTPs however have a chemically cleavable 3'-OH fluorescent reporter group, which only allows the addition of one nucleotide at a time.[308,309] After each addition of such a nucleotide, a camera takes a picture and, depending on the color of the fluorophore, the inserted nucleotide can be determined. The last step consists in the chemical cleavage of the reporter group and the sequencing process can start from the beginning. The stepwise sequencing-cycles make this technology very reliable to detect single nucleotide changes (Figure 45b). Moreover, the possibility to immobilize the primers on a chip allows massive parallel sequencing. Illumina sequencing is significantly superior to Sanger sequencing in terms of the amount of reads possible (100'000 – 100 Mio. reads per run). However, the drawback is the rather inflexible and very short read length (between 50-300 bp).

Since the libraries we envisioned were relatively big (up to 0.5 Mio, spanning ~700 bp), we investigated the Illumina sequencing technology to achieve maximum coverage. Primers for PCR amplification were designed in a way that the region of interest (mutated positions, red) was covered and the PCR amplified fragment was relatively short. The forward primer (blue) was designed to have a complementary sequence for one adapter of the bridge amplification system, and the reverse primer (violet), to have the complementary sequence for the second adapter of the bridge amplification (see Appendix C, Table S19 for primer details). Combining this with a paired end reading, where the fragment is sequenced from both sides simultaneously, sequencing longer fragments was possible. In our case, the Sav gene was PCR amplified to obtain fragments with ~250 bp in size (Figure 45c), and the single chain dimeric version of Sav (scdSav) was PCR amplified to obtain fragments with ~800 bp in length (Figure 45d). In the case of the short fragment, Illumina sequencing led to improved sequencing efficiencies and the sorted library could be analyzed well (Chapter 4.5.3.2). However, for the longer fragment, this sequencing technique was not favorable, leading to fragmented reads (see Chapter 4.4.3 for details).

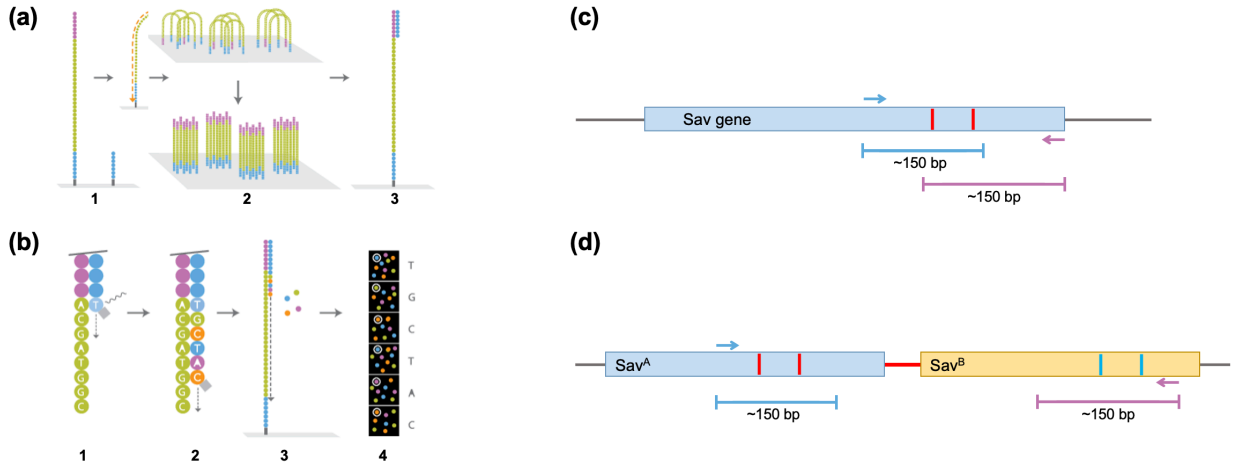


Figure 45. Schematic representation of Illumina sequencing. a) DNA fragments bearing a complementary sequence to the immobilized adapters attach to the chip (1) and are amplified *via* bridge-PCR to form small local DNA colonies (2). The sequencing primer attaches and starts the sequencing process (3). **b)** Sequencing starts by the addition of the first modified nucleotide bearing a 3' blocker (1).[310] After imaging of the chip via a camera (4), the blocker is removed chemically and a next nucleotide is added (2-3). The process is repeated until the defined number of cycles are reached (between 50-300 possible). **c)** Schematic representation of the primer design for the 400-variant library on Sav. Paired end sequencing using 150 cycles, covers both mutated positions (red) from either side. **d)** Schematic representation of the primer design for the 160'000-variant library on scdSav. Paired end sequencing using 150 cycles, covers two mutated positions (red and blue) on either side. The location of each local DNA colony on-chip makes analysis of all four positions simultaneously possible.

4.3.7.3 Nanopore Sequencing

As a final method, Nanopore sequencing was investigated. Nanopore sequencing is a third-generation approach of sequencing especially used to sequence long DNA strands without the need for PCR amplification or chemical labeling. Because of the display of results in real-time and the

comparatively small size of the equipment, the technique found widespread field-use for example in malaria sequencing,[311] on the International Space Station (ISS),[312] and, more recently, worldwide during the Covid-19 pandemic.[313]

The sequencing is based on 'stochastic sensing' which is the determination of the concentration and identity of a substance by analyzing fluctuations in the ionic current. The system used here, is a so called biological nanopore sequencing based on *α*-hemolysin (*α*HL). *α*HL is a 10 nm long transmembrane protein which forms *β*-barrel pore that is wide enough to translocate DNA. The protein is immobilized on a lipid membrane to which a voltage can be applied and that is surrounded by an electrolyte solution. DNA loaded onto this membrane attaches to the pore and translocates through the pore (Figure 46). While doing so, depending on the DNA sequence, i.e. which nucleotide is passing through the pore, the current density across that pore changes and allows for sequence analysis.[314,315]

Here, the Sav gene was amplified using primers that annealed to the plasmid well before and after the Sav gene to obtain long fragments. PCR amplification led to a fragment of 1500-2000 bp in size. Following PCR amplification, the samples were barcoded, the adapter protein attached *via* a blunt/TA ligation, purified over magnetic beads and finally sequenced on a MinION system. For long reads such as in the case of a scanning library (Chapter 4.4.2) or a big 160'00 variants library spanning a longer region of interest (Chapter 4.5) nanopore sequencing turned out to be effective, in spite of the lower number of reads and higher error rates compared to Sanger or Illumina sequencing.

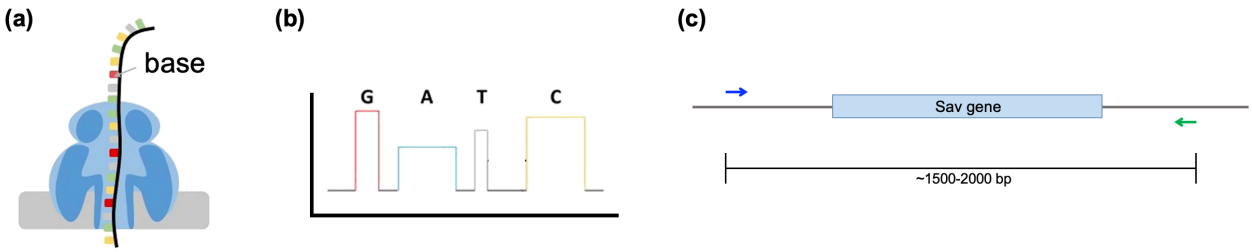


Figure 46. Schematic representation of Nanopore sequencing. a) Immobilized *α*HL pore immobilized on a lipid membrane with a translocating DNA strand.[316] **b)** Different patterns in the fluctuation of the current allows the identification of the corresponding base. **c)** Primer design for the PCR amplification of the Sav gene to obtain long fragments.

4.4 Initial Applications

As discussed in Chapter 3, surface-display of Sav does not afford a homotetrameric quaternary structure. Based on past experience with periplasmic expression of Sav in Top10 (DE3), this compartmentalization strategy was preferred. Periplasmic expression of Sav has been successfully implemented for genetic optimization of ring-closing metathesis,[187] deallylation, hydroamination

and hydroarylation.[317] Top 10 (DE3) was selected as the strain as this allows more efficient library building compared to BL21 (DE3). Importantly the overexpression of Sav using this construct is excellent. (see Appendix C, Figure S31 for expression levels determined by SDS-PAGE).[318]

4.4.1 Screening of Two Known Sav-Variants

First validation experiments involving periplasmic expressed Sav-wt and Sav-MR were conducted to estimate the robustness of the technique. Different ratios of Sav-wt and Sav-MR both labeled with mNectarine, were screened in DEs (Figure 47b). The cells were grown separately in 25 mL shaking flasks for 1-2 h at 37 °C to an OD_{600} = 0.5-0.8 before induction via the addition of IPTG (50 µM final concentration). Expression was carried out overnight (see Appendix C, Figure S31 for expression levels). Individual samples (1 mL, OD_{600}=0.20) with the desired ratios of Sav-wt *vs.* Sav-MR were prepared, centrifuged, the supernatant discarded and the pellet resuspended in PBS (990 µL, pH 7.4). To this cell suspension the cofactor **46** (10 µL, 1 mM, 10 µM final concentration) was added to afford the cells-cofactor mixture for the droplet production. After incubation for 30 min at room temperature, the sample was encapsulated together with substrate **44**. After confirmation of the successful encapsulation and determination of the cell percentage (if the population with mNectarine fluorescence was higher than 30%, the sample was discarded), the sample was incubated at 37 °C for 2-3 h before FACS sorting. The sample was sorted according to increasing coumarin fluorescence and was gated into max. four different gates based on the coumarin fluorescence intensity. Four gates with 10%, ~70%, 10%, and 5% or 1% were sampled, whereby the highest gate was set to 5% or 1% (Figure 47c).

For sorting experiments, a sorting sample consisting of 10 µL of DE solution resuspended in 600 µL of OA (PBS with 0.5% SDS) was prepared, mixed manually and sorted using a BD FACSAria™ II cell sorter. Because of the fluorinated oil forming the oil shell of the DEs, the DEs tend to settle at the bottom of the sample very quickly. To ensure that this does not lead to the loading of too highly concentrated samples in the FACS, the sample tubes were regularly agitated during the experiment. The population was sorted until at least 5% of the corresponding library size was sorted in the highest gate: i.e. for a library size of 1'000 = 50 events in the highest gate, for a library size of 0.5 Mio = 25'000 events. However, we observed that a minimum of 18'000 events was necessary for a successful PCR amplification of the extracted DNA to enable downstream analysis. On average, DE production of ~45 min was sufficient to achieve these numbers in the 5% and/or 1% gate, and the average sorting time was 3-4 h.

To determine the Sav-wt *vs.* Sav-MR ratio, after plasmid extraction, sequencing was carried out to evaluate the enrichment of Sav-MR in the top 5% gate. All three sequencing approaches

confirmed the enrichment of Sav-MR in the highest gate. However, in the case of TOPO cloning, only 10-20 colonies per analyzed gate could be analyzed in total. Nanopore gave an average of 150 reads/gate and NGS had an average of 1500 reads/gate. Sequence analysis revealed a 1.8-fold enrichment of Sav-MR over the unsorted sample with TOPO-cloning, a 1.8-fold enrichment with Nanopore sequencing, and a 3.6-fold enrichment with NGS, for the starting ratio of 1:1 Sav-wt *vs.* Sav-MR (Figure 47d). The difference in the number of reads might be attributed to the variability in enrichment between the different sorts. As NGS provided more reads/gate, and with much bigger libraries in mind, NGS was the preferred sequencing option for the following experiments. Similar enrichment experiments were carried out with different Sav-wt *vs.* Sav-MR ratios, always leading to a significant enrichment of Sav-MR over Sav-wt after catalysis and sorting, and thereby clearly demonstrating the reliability of the assay (Appendix C, Table S8).

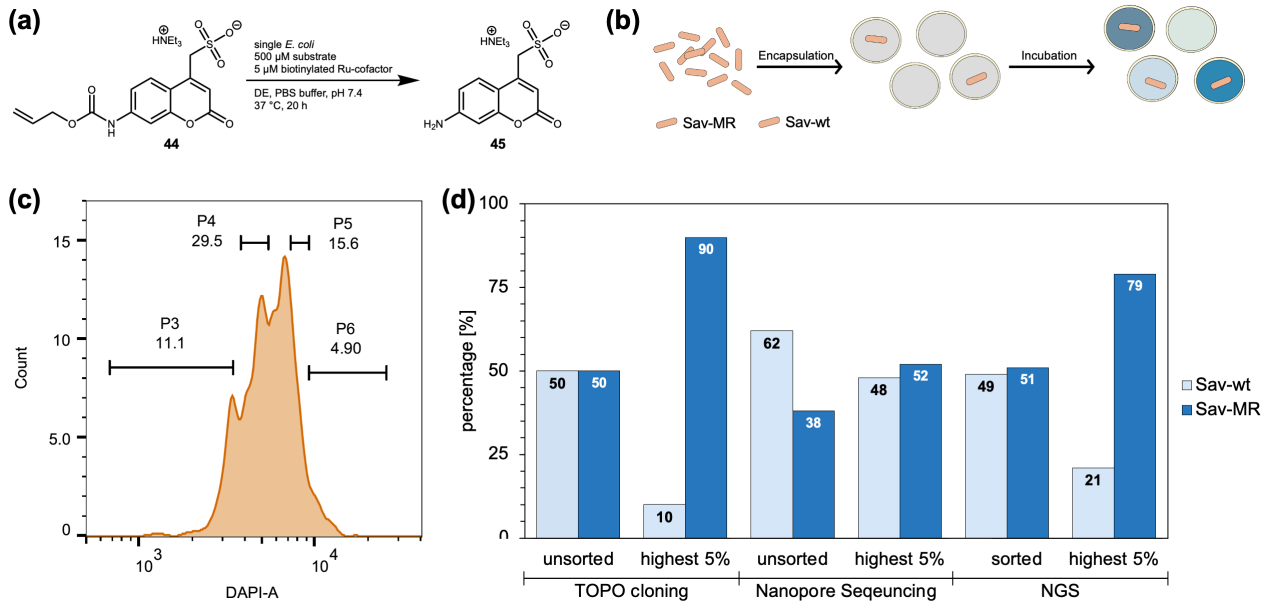


Figure 47. Validation experiment. a) Reaction scheme and conditions for the deallylation of the caged allyl carbamate-aminocoumarin **44** to the aminocoumarin **45** by an ADAse screened in DEs. **b)** mNectarine labeled *E. coli* cells expressing Sav-wt and Sav-MR are co-encapsulated with cofactor **46** and substrate **44** in a 1:1 ratio. **c)** Coumarin fluorescence intensity distribution of the DE population and FACS sorting gates P3-P6. **d)** Comparison of the Sav-wt *vs.* Sav-MR ratio in the top 5% of the population determined by Sanger sequencing, Nanopore sequencing and NGS.

4.4.2 Scanning Library of Streptavidin

In parallel, a scanning library on the whole Sav gene was analyzed. Site saturation mutagenesis was applied to all positions in the Sav gene, yielding 160 x 20 = 3200 variants. The library was bought from Twist Biosciences and incorporated into the periplasmic expression system of Sav on a pET30b vector by Gibson cloning.

Keeping cooperative effects in mind, the aim was to identify hot-spots within the whole gene, including positions far away from the active site, that could have a beneficial impact on catalysis. Preliminary results were obtained by encapsulation of *E. coli* cells expressing periplasmic Sav in

DEs, sorting by FACS and sequence analysis by Nanopore sequencing. Figure 48a displays the sequence analysis of the population with the top 5% coumarin fluorescence intensity. The mutations with the highest occurrence are depicted: one of the most prominent variants was arginine K121R. Even though it was not among the most prominent hits, tyrosine was found to be the most substituted variant at position 112. Sav-YR was previously found to increase the catalytic activity. Moreover, arginine at position 121 was reportedly a very beneficial mutation as observed by Vornholt *et al.*[318] Other potential hotspots include positions 87/88, 120/121 and 144/145. Keeping the fitness landscape of directed evolution in mind, this is a first indication of potentially advantageous, new directed evolution sites for Sav. However, throughout all gates and also in the parent library only an average of 5 mutants were observed at each position, where theoretically 20 are expected (Figure 48b). Therefore, the results need to be considered carefully and optimization of the library preparation is paramount.

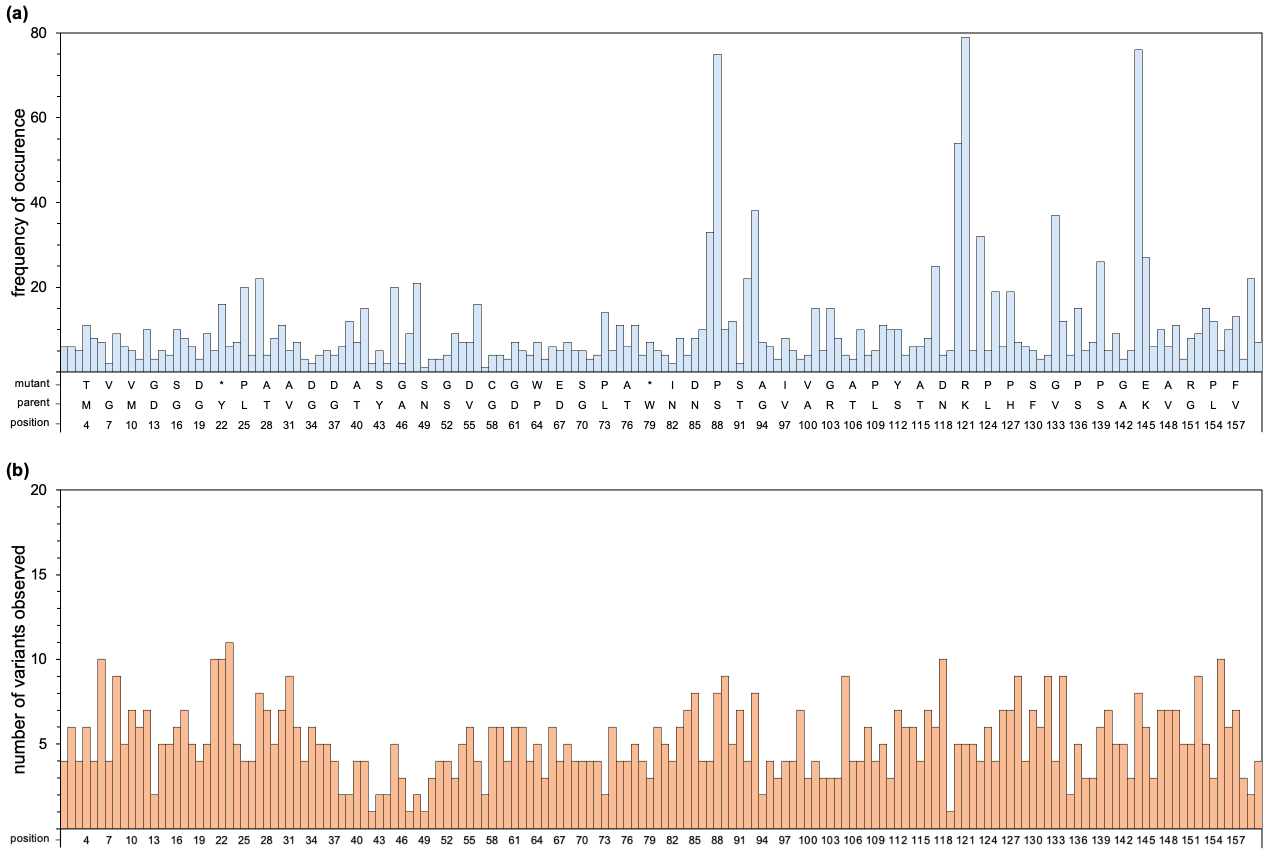


Figure 48. Scanning library screening of Sav. a) Displayed are the variants that have the highest occurrence at the given position. **b)** Displayed is the number of different mutations observed at a given position.

4.4.3 Single-Chain Dimer of Sav

Since we showed that the homotetrameric protein is not displayed as tetramer, the single chain dimeric version of Sav (scdSav) was selected as an alternative.

Using scdSav, it was believed that the dimer will form on the surface of *E. coli* (Figure 49a). This would create a complete active site consisting of both adjacent monomers. However, closer

examination of the scdSav crystal structure (PDB: 6S50) revealed that the diagonally opposite monomers were linked together rather than adjacent monomers (Figure 49b).[190] For this scaffold, periplasmic expression was also preferred.

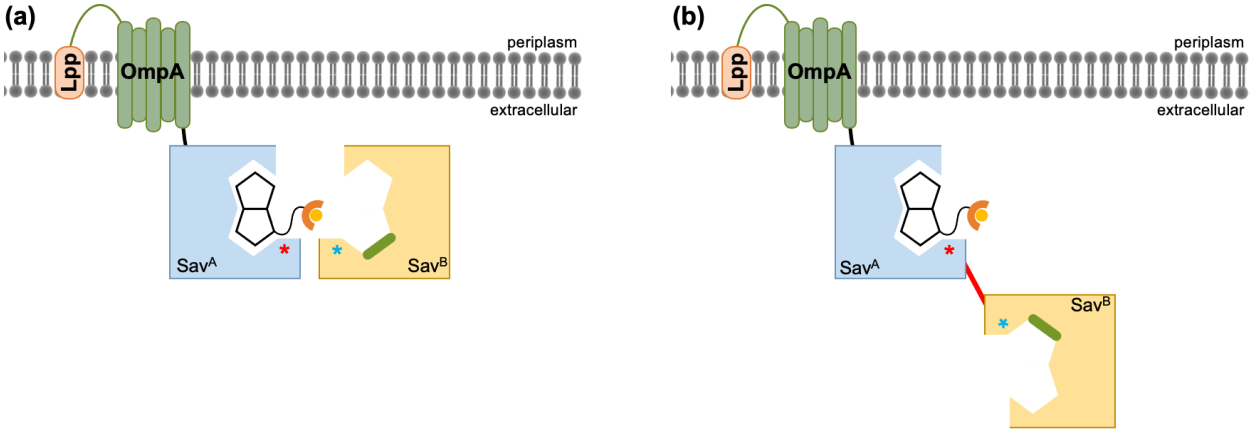


Figure 49. Schematic representation of surface-displayed scdSav using the Lpp-OmpA display system. a) Expected and preferred orientation of the two monomers for the creation of the biotin binding vestibule. **b)** Most probable orientation of both linked monomers according to the crystal structure.[190]

The periplasmic expression of the scdSav-mv2 variant in Top10 (DE3) was investigated. A quadruple mutant library at the positions $S112^A$, $K121^A$, $S112^B$ and $K121^B$ was devised. Initial library construction was carried out by a modified golden gate approach using degenerate codons (Appendix C, Table S17) to introduce the library sequentially first in monomer B and then in monomer A. To do so, a pET30b vector and the Top10 (DE3) strain were used. However, library creation was not successful for one main reason: whenever the second monomer was targeted for mutagenesis, transformation and sequence analysis lead to fragmented sequencing results.

It turned out that the sequence homology between the two monomers was too high (68%), leading to recombination or skips during PCR amplification (see Appendix C, Table S10 for detailed description of skipped regions). Thus, the library was purchased from Twist Biosciences as a combinatorial variant library and incorporated into the pET30b plasmid by Gibson assembly. This yielded the library with a good coverage and, importantly, no bias for a particular mutant. However, some fragmented reads (~5%) and a scdSav-parent background, (~10%) was observed. To compensate for this background, $> 6.0 \times 10^5$ variants were screened.

During sequence analysis after the screening and sorting of the library, the problems with fragmented sequencing reads reappeared. In a small scale NGS run (MiSeq Nano 300) 1.0×10^6 reads were expected. However only ~30'000 reads were generated in the first place and only ~200 reads corresponded to the correct target gene sequence (Figure 50). Closer analysis of individual reads revealed a similar fragmented pattern to the one observed before, but this time appearing in different regions of the gene. Most reads either skipped from one monomer to the other monomer

or only sequenced monomer A/B (Figure 50b). A small set of sequences obtained by Sanger sequencing of directly transformed sorted sample into electrocompetent cells did not have these issues. Moreover, the PCR amplified and purified sample for NGS sequencing showed high purity and only one fragment with an expected size of ~800 bp. Therefore, for the time being, recombination during the expression is unlikely, but not excluded and needs to be further investigated. Because of the way Illumina sequencing works, the sequencing favors short fragments which could explain the observed high ratio for the short and fragmented reads. Just a small fraction of reads corresponds to the expected sequence with all four positions covered (Table 8, entry 1). The majority of reads do either a "type A" skip leading to fragments without any of the four sites (Table 8, entry 2), or only sequence monomer B (Table 8, entry 7). To solve this issue and achieve more manageable library creation options and high-throughput sequencing, the parent scdSav gene sequence needs to be further optimized to lower the sequence homology between the two monomers.

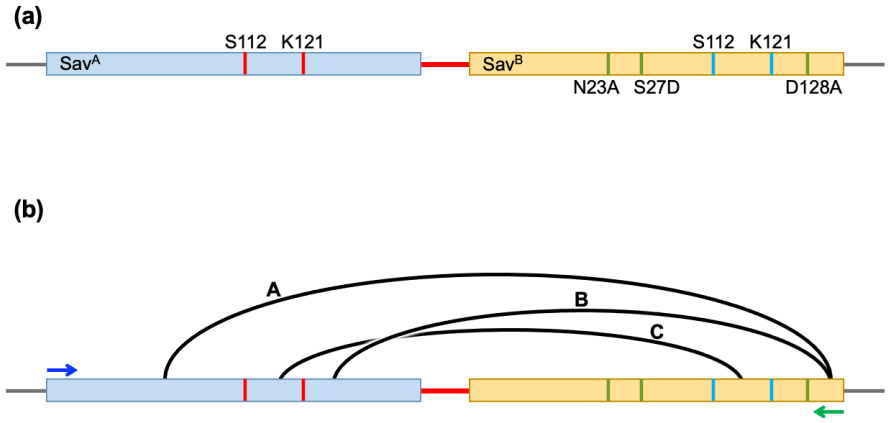
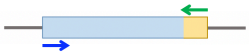
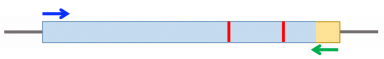
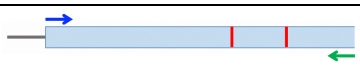
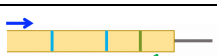


Figure 50. Schematic representation of the sequencing issue. a) Schematic representation of the scdSav gene construct. **b)** Three main sequence skipping patterns identified.

Table 8. Distribution and details of the fragmented reads observed by NGS.

Entry	observed fragment	skip type	occurrence (number of reads)	target monomer	
				A	B
1		no skip	260	✓	✓
2		A	10'000	–	–
3		B	400	✓	–
4		C	7500	~	~
5		no skip	9000	✓	–
6		no skip	7500	–	✓

Data obtained by NGS (MiSeq Nano 300, PE 150). Blue arrows represent the forward primer binding, and forward read. Green arrows represent the reverse primer binding and reverse read. ~30'000 reads were obtained in total and detailed sequence analysis was done using clustal alignment.

Because of the above reasons, nanopore sequencing, primarily developed for long reads, was used as an alternative and the summary of the scdSav screening including the newly found hits, can be found in the following chapter.

4.4.3.1 Application of the DE Screening to a 160'000-variants library.

We then applied the method to screen a library of another isoform of Sav, the single chain dimeric Sav (scdSav).[190] scdSav is a dimeric version of Sav where two adjacent Sav monomers are encoded independently. This was achieved by fusing two subunits of Sav, Sav^A and Sav^B, via a 26 amino acid linker (Figure 51a). The sequence homology of the monomers was minimized to enable the independent mutagenesis of Sav^A and Sav^B. In addition to Sav^B, H127 was mutated to a cysteine to favor the formation of a disulfide bridge in order to form a tetramer quaternary structure. Moreover, the positions N23, S27 and D128 were mutated to alanine, aspartate and alanine respectively in Sav^B only, to obliterate biotin binding in Sav^B. This affords the scdSav isoform that can only bind two equivalents of biotin (instead of the usual four biotin binding sites present in homotetrameric Sav). Most importantly, the two adjacent monomers which constitute the biotin-binding vestibule harboring the cofactor can be mutated independently from each other (Figure 51a). As highlighted in the X-ray structure of [CpRu(QA-Biot)(H_2O)] **46** · Sav-MA (Figure 51b), all four residues Sav^A S112M and K121A and Sav^B S112M and K121A point towards the Ru-cofactor **46**. Mutating the two residues $112^{A/B}$ (and for that matter $121^{A/B}$) independently is not possible when using the homotetrameric Sav.

Accordingly, we targeted and randomized simultaneously positions $S112^A$, $K121^A$, $S112^B$ and $K121^B$ for mutagenesis.[223,318] A site saturation library at all four positions was incorporated leading to a theoretical library size of 1.6×10^5. Stop codons that could lead to lower diversity of the library were not designed and incorporated and no bias for one particular amino acid was found. However, a ~10% bias towards the parent scdSav, and another ~10% background (fragmented sequences, frame shifts, stop codons) were observed and was compensated by screening > 0.5 Mio variants to cover all possible variants with a 95% probability.

The library was screened by encapsulating mNectarine-labeled *E. coli* cells with plasmids for the expression of the scdSav-library in DEs together with the substrate **44** and cofactor **46** (Figure 51c). The collected DEs were incubated at 37 ºC and their fluorescence was analyzed at different time points using a flow cytometer (Figure 51d). We observed that the coumarin FI increases over the first 3.5 h of incubation, while longer incubation times did not lead to a substantial increase in fluorescence intensity. Accordingly, DE samples were incubated for 4 h at 37 °C and then sorted by

FACS into four gates with increasing aminocoumarin FI (Figure 51e). A fifth gate containing the top 1% of DEs with highest activity was collected.

Sequencing results obtained by Nanopore sequencing, revealed the increased occurrence of the variant with mutations S112T and K121G in monomer A, and S112M and K121S in monomer B (scdSav-TGMS) in the top 1% gate. Two further prominent variants were scdSav-WYKK and scdSav-LMMS. Moreover, it should be noted that S112M and K121S in monomer B were prominent mutations throughout, and were with 33 times even more often represented as the best quadruple hit (see Appendix C, Table S12 for summary of Nanopore sequencing results). *In vitro* activity assays are currently being carried out to evaluate the effective improvement in catalysis rates over WT. To evaluate and interpret the new quadruple mutations, we modeled the mutations into a crystal structure of cofactor **46** · Sav (PDB: 6FH8) by PyMol (Figure 51g).

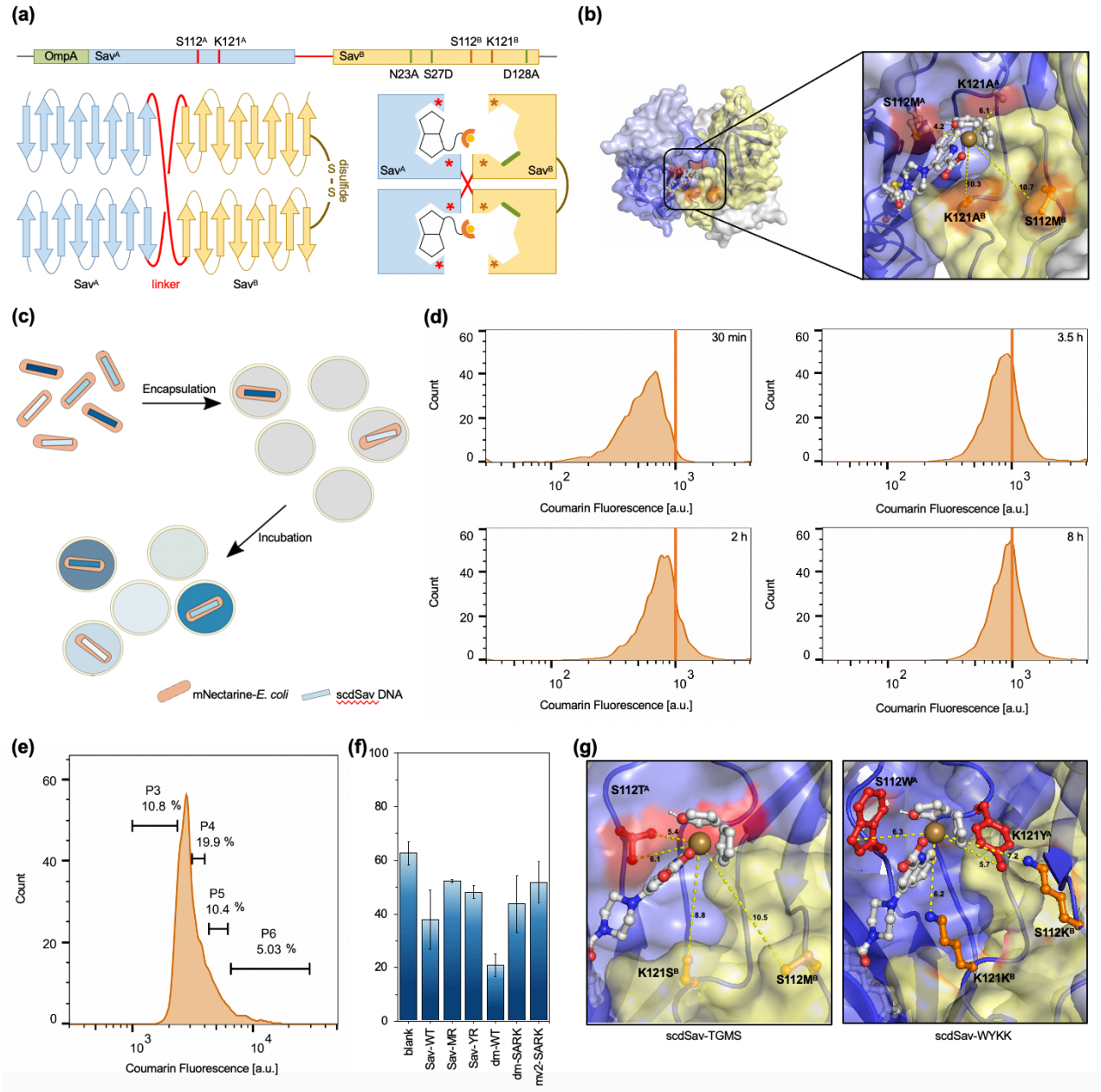


Figure 51. Screening of the scdSav-library. a) Schematic representation of the scdSav coding region of the pET30b vector. The sequence homology of monomer A (Sav^A, blue) and monomer B (Sav^B, yellow) is reduced and the two monomers are linked via a 26 amino acid linker (red). Residue H127 is mutated to H127C in Sav^B to favor the formation of a disulfide bond (gold line). Three additional mutations to N23A, S27D and D128A are introduced (dark green) to obliterate the biotin binding in Sav^B. Secondary structure of the scdSav tetramer and quaternary structure of the scdSav with two incorporated biotinylated metallocofactors (blue = Sav^A, yellow = Sav^A, blue and yellow stars = mutated positions, black pentagons = biotin, orange = ligand, yellow = Ru core). **b)** Crystal structure of the ADAse [CpRu(QA-Biot)(H_2O)] **46** · Sav S112M-K121A (PDB-ID: 6FH8). The four units are displayed as transparent surface in blue, pale blue, yellow and pale yellow. The mutated positions in Sav^A ($S112M^A$ and $K121A^A$, red) and the positions facing the active site in Sav^B ($S112M^B$ and $K121A^B$, orange) are highlighted and labelled. The biotinylated ruthenium complex **46** is displayed as ball and stick. Color code: C = grey, N = blue, O = red, S = yellow, H = white, Ru = dark yellow sphere. The cofactor **46** is anchored in the Sav monomer A (blue). **c)** mNectarine-labeled *E. coli* expressing scdSav are encapsulated together with cofactor **46** and substrate **44**. **d)** Coumarin fluorescence intensity distribution of DEs encapsulating the scdSav-library at 30 min, 2 h, 3.5 h and 8 h after encapsulation. The red bar at 10^3 is there for a comparison purpose. **e)** Fluorescence intensity distribution of DEs containing the scdSav-library with the sorting gates P3-P6 from low intensity to high intensity: P3 = 10%, P4 = 50%, P5 = 10% and P6 = 5%. **f)** Catalysis in microtiter plates for selected variants. **g)** Structure

obtained by modeling the quadruple mutants into an available crystal structure with incorporated cofactor **46** (PDB-ID: 6FH8) by PyMol. The four units are displayed as surface in blue, pale blue, yellow and pale yellow. The mutated positions in Sav^{A} are displayed in red and the positions facing the active site in Sav^{B} are highlighted in orange. The biotinylated ruthenium complex is shown as balls and sticks. Color code: C = grey, N = blue, O = red, S = yellow, H = white, Ru = dark yellow sphere. The cofactor **46** is anchored in the Sav monomer A (blue).

In the PyMol model we can see, that in the case of scdSav-TGMS mostly polar coordinating residues were introduced. The higher catalytic activity might arise from a possible tighter retention of the metal cofactor within the active site by residues such as threonine and serine. Moreover, the introduction of glycine at position 112 in monomer A, seems to create more space, which might allow better access to the substrate. In contrast, the introduction of the rather hydrophobic residues tryptophan and tyrosine in scdSav-WYKK, might lead to π-π stacking interactions with the quinoline ring of the ligand and the Cp-moiety respectively. This could likewise lead to an overall better stabilization of the cofactor within the pocket. Arginine proved to have beneficial effects on this particular reaction multiple times. Similarly, the introduction of positively charged lysines resembles the appearance of arginine in previously reported studies of the same scaffold. These hypotheses are currently being investigated by *in vitro* screening and crystallography of the newly found quadruple mutants.

The following section has been published in
Angew. Chem. Int. Ed. **2022**, *61,* e2022073.
doi: 10.1002/anie.202207328

Ultrahigh-Throughput Screening of an Artificial Metalloenzyme using Double Emulsions Screening

Jaicy Vallapurackal[#[a]], Ariane Stucki[#[b]], Alexandria Deliz Liang[[a]],

Petra S. Dittrich*[[b]], Thomas R. Ward*[[a]].

[a] University of Basel, Mattenstrasse 24a, CH-4058 Basel, Switzerland. E-mail: thomas.ward@unibas.ch

[b] Department of Biosystems Science and Engineering, ETH Zurich, Mattenstrasse 26, CH-4058 Basel, Switzerland. E-mail: petra.dittrich@bsse.ethz.ch

[#] These authors contributed equally

KEYWORDS: directed evolution · *in vivo* screening · ultrahigh-throughput

4.5.1 ABSTRACT

The potential for ultrahigh-throughput compartmentalization renders droplet microfluidics an attractive tool for the directed evolution of enzymes. Importantly, it ensures maintenance of the phenotype-genotype linkage, enabling reliable identification of improved mutants. Herein, we report an approach for ultrahigh-throughput screening of an artificial metalloenzyme in double emulsion droplets (DEs) using commercially-available fluorescence-activated cell sorters (FACS). This protocol was validated by screening a 400 double-mutant streptavidin library for ruthenium-catalyzed deallylation of an alloc-protected aminocoumarin. The most active variants, identified by next generation sequencing, were in good agreement with hits obtained using a 96-well plate procedure. These findings pave the way for the systematic implementation of FACS for the directed evolution of (artificial) enzymes and will significantly expand the accessibility of ultrahigh-throughput DE screening protocols.

4.5.2 INTRODUCTION

Progress in protein engineering, DNA sequencing, and bioinformatics has enabled the engineering of enzymes to a specific need. While enzyme properties have evolved over thousands of generations by natural selection, directed evolution offers tools to accelerate and streamline this process by screening enzyme libraries and selecting the most promising candidates for a given function. Early directed evolution campaigns have yielded enzymes with increased solvent tolerance[3] or significantly enhanced catalytic activity[13] and selectivity[6–8]. Despite these achievements, the full potential of directed evolution has not yet been realized due to the requirement that the genotype-phenotype linkage be preserved during the entire screening process. Accordingly, most directed evolution campaigns rely on screening libraries in microtiter plates (MTPs). Such screening campaigns are time- and resource-intensive and scale exponentially with the number of amino acid positions screened simultaneously[2,19]. These aspects impose severe limitations on the library size, thus limiting scientists either to the screening of single positions iteratively or only a limited portion of the possible variant landscape.

Within the realm of evolvable enzymes, artificial metalloenzymes (ArMs) have recently attracted increasing attention. Consisting of an abiotic metal cofactor anchored within an evolvable protein scaffold, these hybrid catalysts can exhibit unique new-to-nature reactivities[160]. In this context, ArMs based on the biotin-streptavidin (biot-Sav) technology have been genetically engineered to catalyze numerous reactions including metathesis[187] as well as transfer hydrogenation, hydroamination, and hydroxylation[159]. Other versatile ArM-scaffolds include carbonic anhydrase[195], hemoproteins[192,319,320], prolyl oligopeptidase[321], four-helix bundles[322],

the lactococcal multiresistance regulator[193], or even *de novo* designed metallopeptides[323]. Building on the pioneering work of Meggers and coworkers[210,211,213], streptavidin-based ArMs that catalyze the deallylation of allyl-carbamate-protected coumarin were previously screened in 96-well plates[179,188]. An artificial deallylase (ADAse) contained in the periplasm of *E. coli* cells was optimized by simultaneously randomizing two amino acid residues, to afford a 20-fold improvement in activity in a single round[188]. These results highlighted that (i) ArMs can be evolved to catalyze new-to-nature and bioorthogonal reactions, (ii) the periplasm of an *E. coli* offers a hospitable environment to compartmentalize ArMs while maintaining the phenotype-genotype linkage and (iii) in contrast to iterative saturation mutagenesis, the systematic screening of multiple positions simultaneously enables the identification of potential synergistic mutations. However, the lack of high-throughput screening tools typically restricts these studies to the MTP screening of libraries containing less than 500 mutants.

Advances in microfluidics, and particularly in droplet-based microfluidics, over the past 20 years have led to the development of tools enabling high-throughput screening of large libraries of enzymes[23,324]. Such tools are based on the encapsulation of single genetic variants (single cells or DNA molecules) in aqueous compartments (droplets) together with the reaction components. Monodisperse droplets can be produced on-chip at throughputs of several thousand Hertz, resulting in the encapsulation of large libraries within seconds to minutes. Each droplet, isolated from its surroundings by oil, provides a means of maintaining the phenotype-genotype linkage. Additionally, the use of fluorogenic reactions has enabled further advances in high-throughput sorting—such as on-chip fluorescence-activated droplet sorting—and has led to the optimization of various enzymes[23,98,99,101,102,324,325]. However, even though recent publications have reported on high throughput droplet sorting[98], the droplet sorting speed of most microfluidic platforms is typically significantly lower (~a few Hz)[326] than the achievable production rate (~several kHz)[73], and decreases even further if the sample is sorted into more than two distinct populations. Moreover, on-chip sorting of single emulsion water-in-oil droplets requires custom-engineered chips and software, thus limiting the use of microfluidic screening to highly specialized teams.

The challenges associated with screening of single emulsion droplets may be addressed through the use of double emulsion (water-in-oil-in-water) droplets (DEs), which, as a result of their aqueous exterior, are compatible with commercially-available ultra-fast flow cytometers and FACS available at many academic facilities[143,144,303]. The production of DEs for directed evolution has typically involved the use of batch methods, which lead to polydisperse droplets with multiple inner aqueous phase compartments[113,114,134]. Recent developments, however, have led to the high-throughput on-chip generation of water-in-oil-in-water droplets, analogous to single emulsion

droplets, and their use in the directed evolution of natural enzymes[41,135]. The use of DEs enables compartmentalization of single genetic variants (i.e., genotype) with the corresponding reaction products (i.e., phenotype), and high-throughput sorting of the DEs into several populations can be achieved using FACS. Importantly, the sorting throughput is comparable to the droplet production rate (kHz), thus enabling streamlining and automatization of the entire process. Furthermore, the use of widely available FACS instrumentation renders droplet sorting more accessible to a broader scientific community. Accordingly, this strategy is an attractive tool for high-throughput directed evolution of enzymes. Here, we describe an ultrahigh-throughput assay based on DE microfluidics for the *in vivo* directed evolution of an artificial deallylase (ADAse) based on biot-Sav technology (Figure 52). The method is validated initially by carrying out a model enrichment and further validated by screening a 400 double-mutant library[188]. The rapid screening enabled by this protocol has the potential to transform directed evolution of enzymes. Indeed, while one researcher screening sixteen 96-well plates per week would require approximately six years to screen 500,000 variants, the DE method shortens this workflow to a single week, more than a 300-fold reduction in turnaround time.

4.5.3 Results and Discussion

The catalytic system used in this work is a previously described ArM[179] based on Sav compartmentalized in the periplasm of *E. coli* and the biotinylated ruthenium cofactor 1 (Figure 52)[188]. The *E. coli* are equipped with two different plasmids. The first is a pUA66 vector encoding mNectarine with chloramphenicol (cam) resistance (Figure 52a). The red-fluorescent protein mNectarine is therefore expressed constitutively. The second plasmid is a pET30b vector encoding a T7-tagged Sav fused to the signal peptide of the outer membrane protein A (OmpA) with kanamycin (kan) resistance. Upon addition of isopropyl β-D-1-thiogalactopyranoside (IPTG), Sav expression is induced in the cytoplasm, and the protein is subsequently secreted to the periplasm, where it forms the homotetrameric protein consisting of four *β*-barrels that can bind up to four equivalents of biotin[12] (Figure 52b). Following expression, the cofactor 1 is added and the ADAse self-assembles in the periplasm as a result of the high affinity of Sav for biotin (Figure 52c,d). The allylcarbamate-protected aminocoumarin 2 is deprotected by the catalyst to afford the fluorescent aminocoumarin 3, which can be used to evaluate the catalytic activity of the encapsulated ADAse (Figure 52e).

For the ultrahigh-throughput screening of the ADAse, we utilized the approach presented in Figure 53. In brief, mNectarine-labeled *E. coli* cells expressing Sav were incubated with the

ruthenium cofactor **1** and then encapsulated with an allyl-carbamate protected aminocoumarin substrate **2** (Figure 53b) in DEs on a microfluidic chip. The resulting droplets were collected off-chip, incubated under the desired reaction conditions, and then sorted using FACS based on the fluorescence intensity (FI) of both mNectarine and coumarin (Figure 53c,d). The presence of an mNectarine signal enabled the sorting of DEs containing an *E. coli* cell, while the coumarin fluorescence served as a readout for catalytic activity. After sorting, the DEs were ruptured, the plasmid was extracted, and the gene of interest was PCR-amplified and analyzed by next generation sequencing (NGS) (Figure 53e,f). Note, the hits identified in this way may be subjected to another round of the assay to iteratively evolve the enzyme of interest.

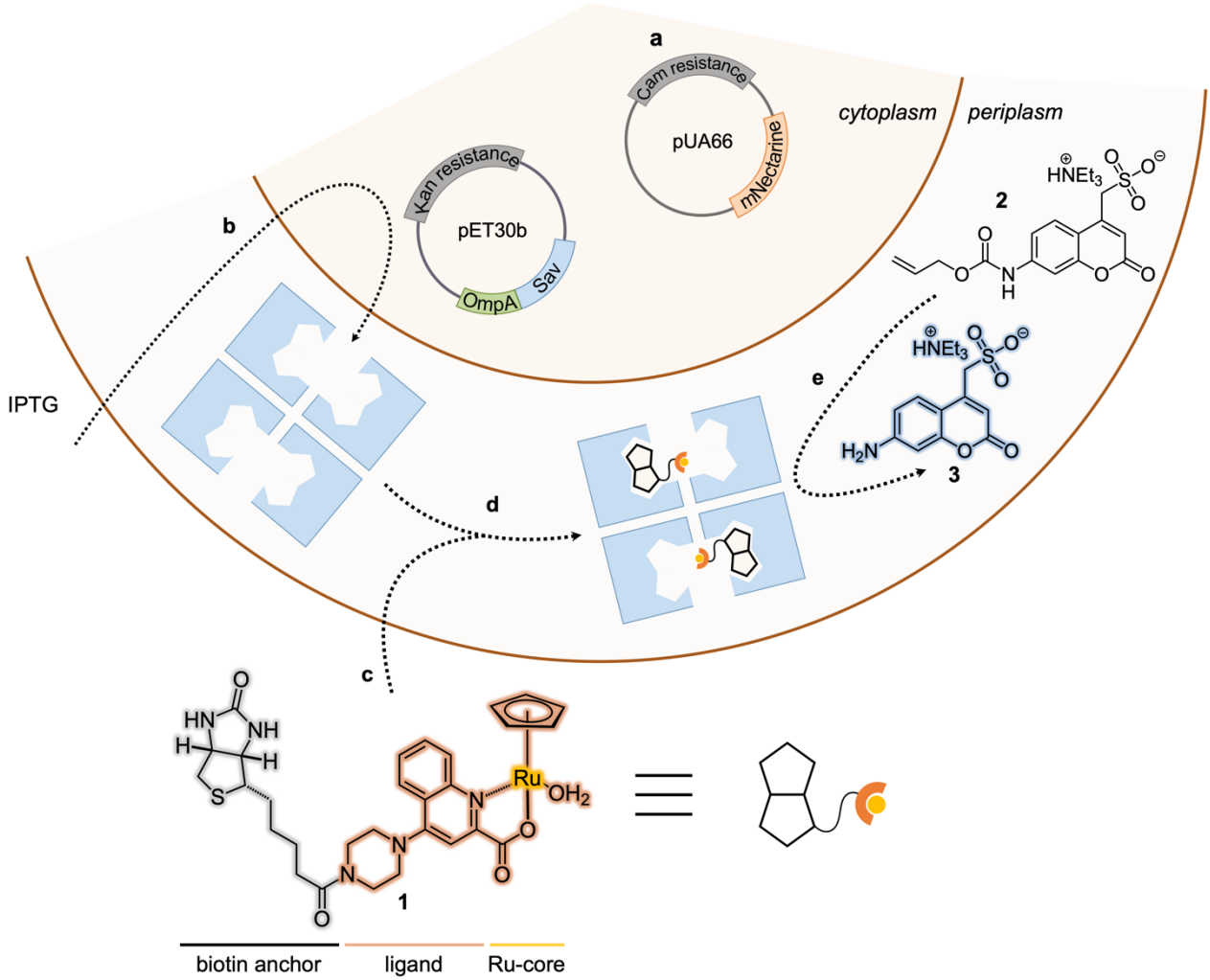


Figure 52. Model system used for the directed evolution of artificial metalloenzymes using a microfluidics-based screening assay. a) mNectarine is encoded on a pUA66 vector and is constitutively expressed and compartmentalized in the cytoplasm, enabling the fluorescence detection of DE droplets containing an E. coli cell. **b)** Sav is encoded on a pET30b vector and induction with IPTG leads to the overexpression and secretion of Sav into the periplasm of E. coli, where it forms a homotetrameric protein. **c)** The biotinylated ruthenium cofactor 1 passively diffuses through the outer-membrane into the periplasm and **d)** is anchored in the binding pocket of Sav, affording the artificial deallylase (ADAse). **e)** Uncaging of the allyl-carbamate protected substrate 2 yields the fluorescent aminocoumarin 3, which can be detected and sorted by FACS.

4.5.3.1 CHARACTERIZATION OF THE DOUBLE EMULSION-BASED SCREENING ASSAY.

Monodisperse DEs with a diameter of about 15 µm were produced on polydimethylsiloxane (PDMS) chips at rates >6000 Hz (see Appendix C, Figure S37 and S38 for details)[303]. With on-chip production, monodisperse droplets (coefficient of variation typically ~3%) are produced, i.e. variations of the volume and any related errors due to polydispersity can be neglected. Initially, DEs with a diameter of 25 µm and a volume of ~8.2 pL were tested. After optimization of the chip design, the final diameter and volume were reduced to 15 µm and ~1.8 pL respectively. This droplet volume is ideal as it allows for quick accumulation of the product and fits perfectly in the FACS workflow. Moreover, since only one *E. coli* is encapsulated per droplet, the smaller volume of the droplet leads to a higher concentration of the ArM. This feature is reflected in an increased conversion and thus a more reliable signal to noise ratio. To produce DEs containing the reaction mixture, the following protocol was applied: a solution of *E. coli* previously incubated with cofactor **1** in phosphate buffered saline (PBS) was introduced into the chip via a first inlet. Since the biotinylated cofactor consists of a CpRu-precursor and a biotinylated ligand (Figure 1), both stock solutions were prepared in dimethylformamide (DMF) to ensure their dissolution and pre-incubated with *E. coli* (expressing Sav in the periplasm) in PBS (+0.5% DMF) (for details, see SI). The high affinity of biotin for Sav (K_d ~ 10^{-14} M) ensures that the biotinylated hydrophobic cofactor remains compartmentalized in the periplasm of the *E. coli* after binding to Sav and does not diffuse out of the droplet. The solution of the substrate **2** in PBS was added via a second inlet. The use of two inner aqueous (IA) phase inlets shields the *E. coli* and ADAse from the substrate **2** until on-chip encapsulation. The PDMS chip, the encapsulation conditions and the dilution of the cell suspension were optimized to afford ~77% empty droplets, ~20% droplets containing one cell, and the remaining 2.6% containing two or more cells per droplet. Although there are reported methods to improve this ratio further[70], the throughput of both droplet production and FACS, allowed us to conveniently work with this ratio.

E. coli can survive in DEs for at least 10 hours, as previously shown by growth experiments and fluorogenic assays[303]. However, the overexpression of Sav leads to premature cell-death, as it depletes the cell from biotin[326]. However, as the phenotype-genotype linkage is maintained within each droplet, cell survival is not required to enable deconvolution of the best Sav mutants by NGS.

After collection and incubation of the DEs off-chip, the fluorescence in the inner aqueous phase was quantified by FACS. The DEs were first gated based on forward- and side-scatter profiles to separate them from oil droplets and from non-homogeneous DEs, e.g., those containing two inner aqueous phase droplets (Figure 54b). Empty DEs could be further sorted based on the absence of a fluorescent signal from mNectarine.

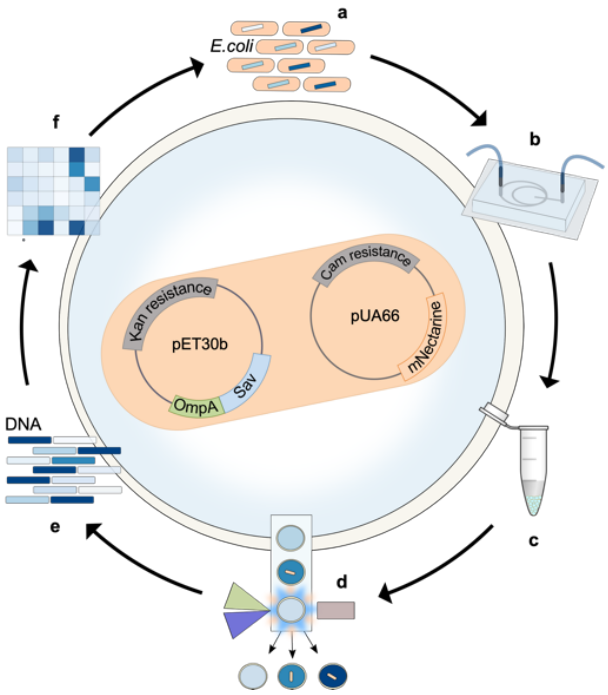


Figure 53. Assay workflow for the screening of ArMs in double emulsion droplets. a) A library of *E. coli* harboring Sav in their periplasm and mNectarine is produced and incubated with cofactor **1** in phosphate-buffered saline. **b)** The E. coli library and the substrate **2** are fed onto a microfluidic chip in two separate solutions and encapsulated together in double emulsion droplets (DEs). **c)** After incubation at 37 °C off-chip, **d)** the DEs are subjected to FACS, which enables a dual-channel sorting to enrich a population of droplets containing an *E. coli* cell (as highlighted by the FI of mNectarine, λ_{ex} = 560 nm, λ_{em} = 580 nm) and displaying high catalytic activity (as revealed by the FI of aminocoumarin 3, λ_{ex} = 405 nm, λ_{em} = 460 nm). **e)** Plasmid extraction of the sorted droplets yields plasmid DNA, which is amplified by PCR to enable **f)** NGS analysis of the sorted library. Hits can be transformed back into E. coli to reiterate the cycle.

To quantify the coumarin **3** FI and stability, different DE populations were prepared with substrate **2** (500 µM) and **3** in three different concentrations (5, 50 and 500 µM). Flow cytometry analysis of these populations one and 24 h after encapsulation confirmed the stability of both the substrate and product, with minimal leakage in the course of 24 h (Figure 54c). The presence of a sulfonate group on both substrate **2** and product **3** minimizes their diffusion into the hydrophobic oil-phase. Furthermore, these results demonstrate that the three product concentrations can be clearly distinguished from the background fluorescence of substrate **2**. We also investigated whether the presence of *E. coli* cells affects the stability of substrate **2** or the catalytic activity of the ADAse. In particular, we co-encapsulated (i) *E. coli* cells harboring wildtype streptavidin (wt-Sav) in their periplasm with substrate **2** (500 µM) and (ii) *E. coli* cells lacking the plasmid for Sav expression (previously incubated with 5 µM of **1**) together with the substrate **2** (500 µM). In both cases, DEs containing cells were compared with DEs without cells after 0 h and 10 h of incubation. In the case of (i), the fluorescence determined for DEs with and without cells was within the same range and remained constant over time (Figure 54d). For scenario (ii), both DE populations again displayed very similar fluorescence intensity initially and similar increases in fluorescence over time as more substrate **2** was converted into product **3** (Figure 54e).

Next, the screening method was validated by conducting an enrichment experiment wherein mNectarine-labeled *E. coli* expressing wt-Sav or a known variant Sav-MR were combined in a 99:1 ratio, incubated with cofactor **1**, and then encapsulated in DEs along with the substrate **2** as described earlier (Figure 55a). The variant Sav-MR bears two mutations at S112M and K121R, which leads to improved ADAse activity[188]. After incubation, the DE sample was sorted by FACS and the population of droplets with highest coumarin FI was collected (Figure 55b). After plasmid extraction, NGS was carried out to determine the enrichment of Sav-MR in the top 5% gate. The enrichment factor was determined by applying the following formula: (Sav-MR$_{top5\%}$ / wt-Sav$_{top5\%}$) / (Sav-MR$_{unsorted}$) / wt-Sav$_{unsorted}$). Sequence analysis revealed a 68-fold enrichment of Sav-MR relative to the unsorted DEs sample, thereby validating the reliability of the assay (Figure 55c).

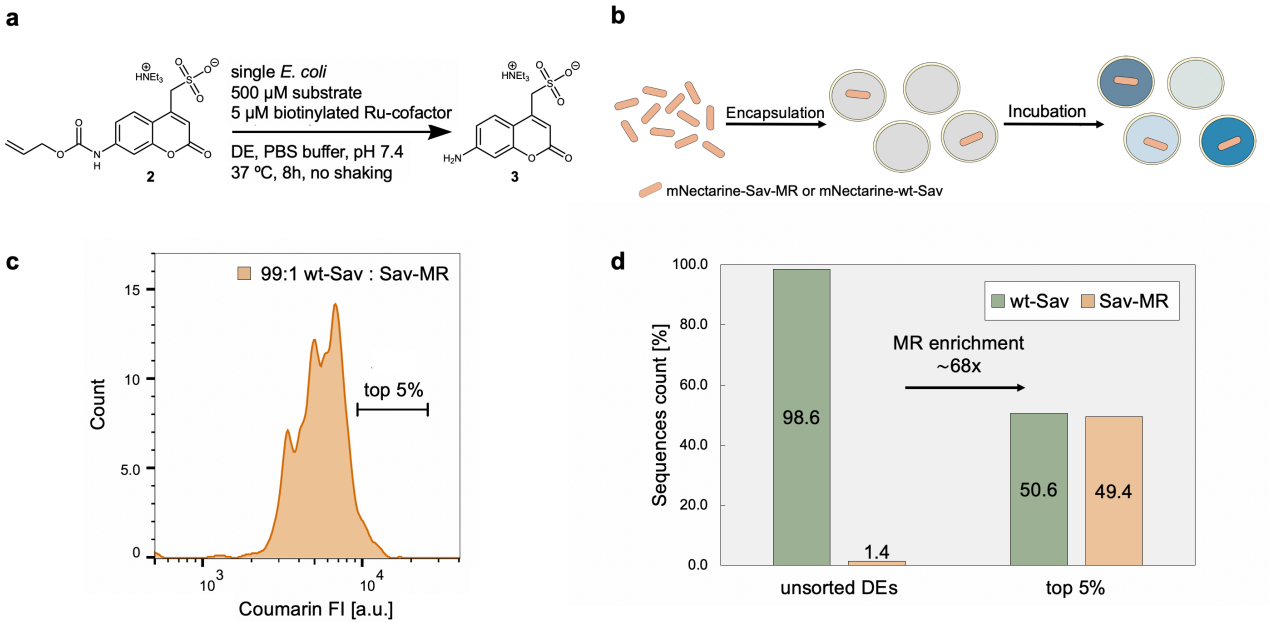


Figure 54. Control experiments with wt-Sav and Sav-MR. a) mNectarine-labeled E. coli expressing wt-Sav and Sav-MR (99:1 ratio) were pre-incubated with cofactor 1 (5 µM) in PBS buffer, encapsulated in DEs with substrate 2 (500 µM), and incubated at 37 °C for 8 h with no shaking. **b)** The coumarin FI distribution of the resulting DEs and the top 5% gate selected for analysis by FACS. **c)** Enrichment of Sav-MR with respect to wt-Sav in the top 5% of the population determined by NGS.

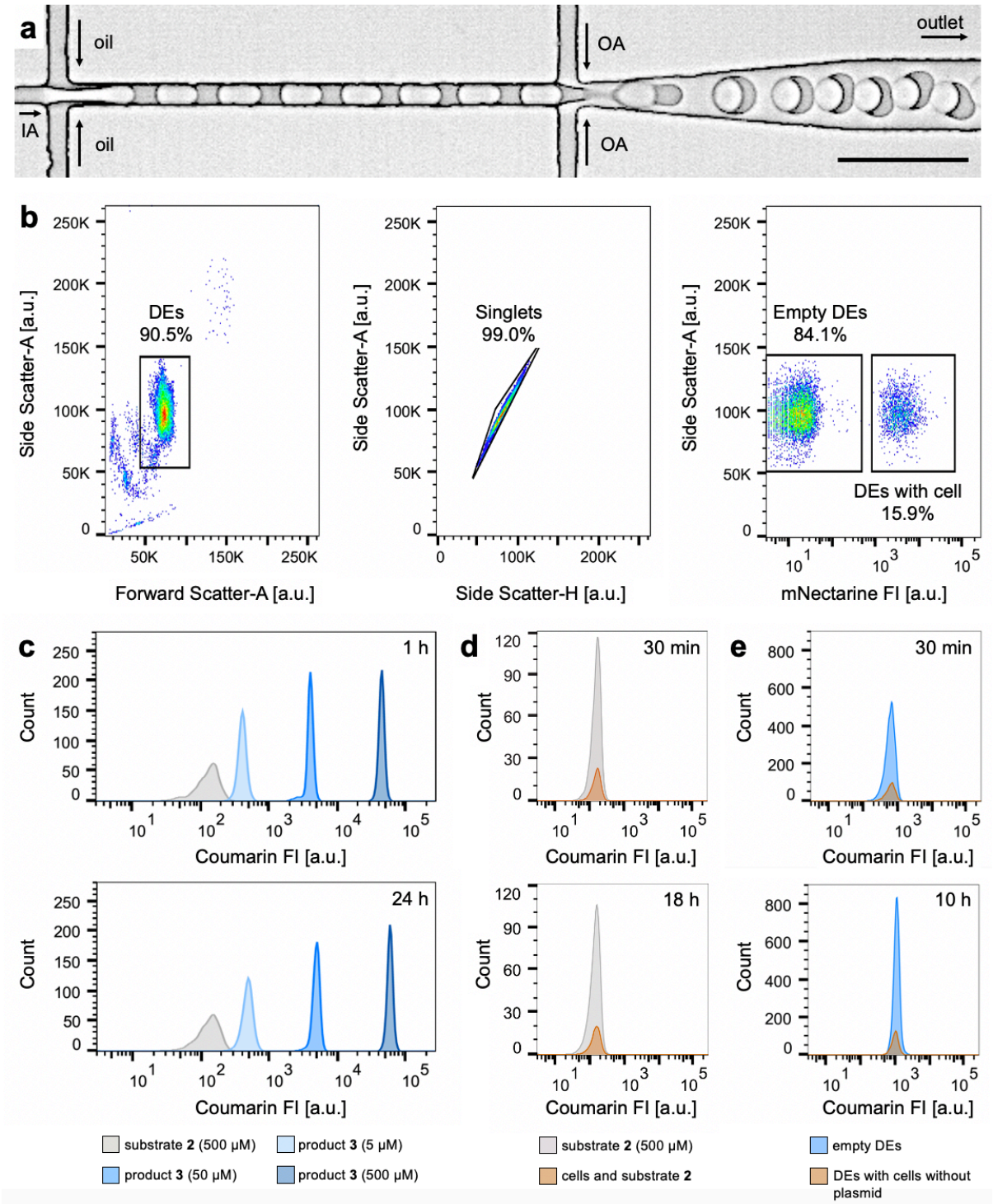
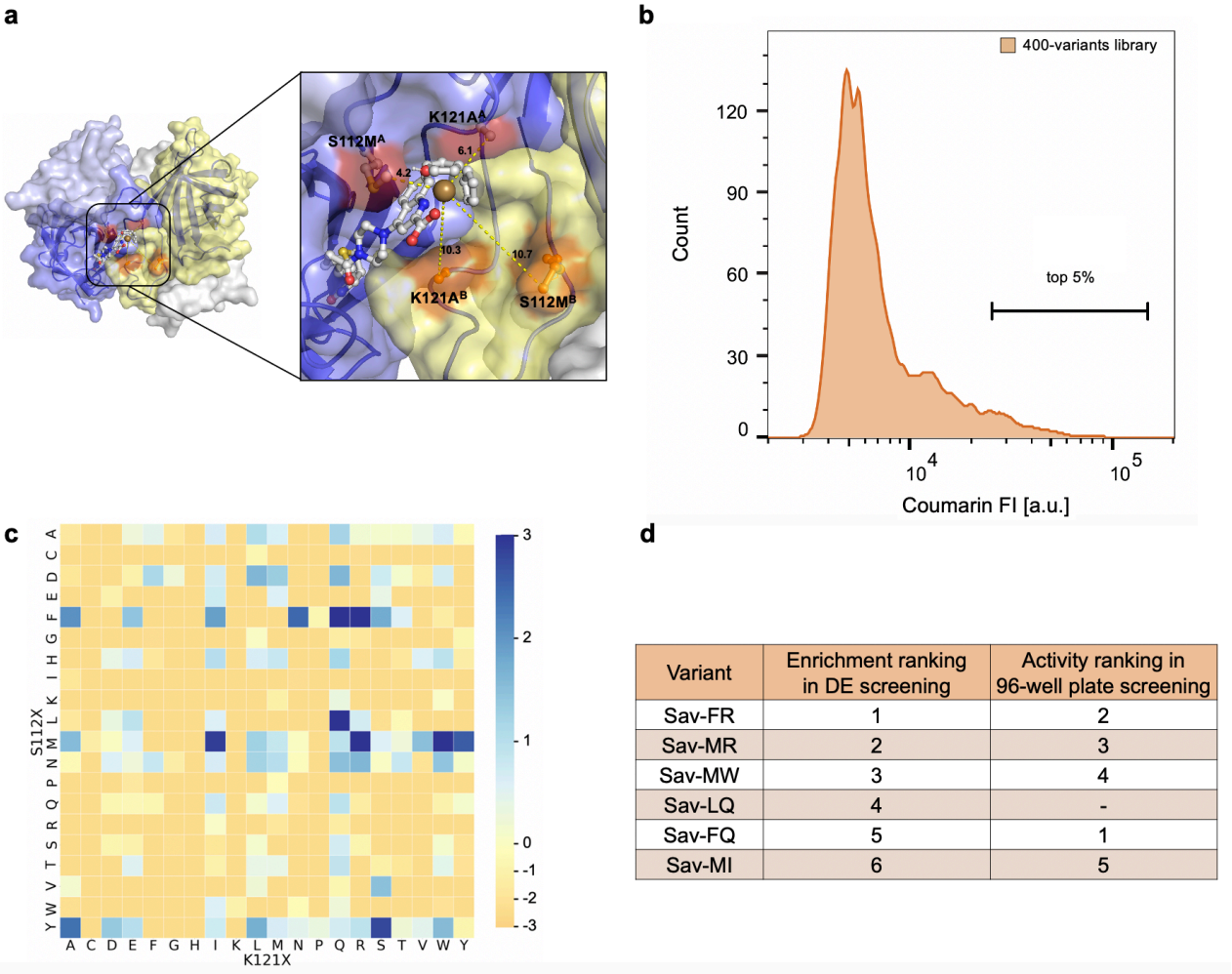


Figure 55. DE formation and reaction monitoring by FACS.a) Micrograph of DE formation. Monodisperse DEs co-encapsulating E. coli cells, cofactor 1 and substrate 2 are produced using a PDMS-based microfluidic device. IA = inner aqueous phase, OA = outer aqueous phase, oil = oil phase, scale bar: 50 µm. **b)** Left: Flow cytometer light scatter gate of a DE sample displaying 10,000 events randomly sampled, revealing the size and shape distribution of the sample. In this case ~91% of the sampled DEs are monodisperse. Middle: Flow cytometer light scatter gates of homogeneous DEs from the selected DE subpopulation. This plot was used to distinguish DEs with single aqueous cores from DEs with multiple aqueous cores. Right: mNectarine FI of the selected DE subpopulation enabling the sorting of DEs containing an E. coli cell and empty DEs. **c)** Coumarin FI distribution of four different DE samples, encapsulating respectively 500 µM substrate 2 or different product 3 concentrations (5, 50 and 500 µM), analyzed 1 h (top) and 24 h (bottom) after encapsulation. **d)** Coumarin FI distribution of DEs encapsulating 500 µM substrate 2 and E. coli featuring wt-Sav, analyzed 30 min and 18 h after encapsulation, highlighting that substrate 2 is inert in the absence of the cofactor 1. **e)** Coumarin FI distribution of DEs encapsulating substrate 2 (500 µM) and cells without plasmid for Sav expression (pre-incubated with 5 µM of cofactor 1), obtained 30 min and 10 h after encapsulation.

4.5.3.2 VALIDATION OF THE ASSAY BY THE SCREENING OF A 400-VARIANT LIBRARY.

The method was next applied to the screening of a 400-variant library of Sav double mutants bearing mutations at positions S112 and K121. This library was previously screened using an automated 96-well plate assay, which revealed six variants with activities ≥ 12-fold higher than wt-Sav for the deallylation of **2**, namely Sav-FQ, -FR, -MR, -MW, -MI and -AW (Figure 56b, Appendix C, Tables S13 and S14)[188]. After preparation of the DE sample and incubation, the top 5% of the DEs with highest coumarin FI were sorted by FACS (Figure 56b). Plasmids were extracted from the sorted and unsorted populations and analyzed by NGS to determine the enrichment factor of the 5% gate. Notably, of the six variants previously identified using the automated 96-well plate assay, five (excluding Sav-AW) were identified by NGS as having the highest enrichment in the top 5% gate (Figure 56c and Figure S38). Additional variants including Sav-LQ and Sav-MY were also enriched in the top 5% gate relative to the unsorted sample. We confirmed the reproducibility of these results by repeating the screening using a biological replicate. A comparison of the 20 most enriched variants determined from both data sets revealed an overlap of 75% and a good correlation for the enrichment of all mutants (see Appendix C, Figure S39, $R^2 = 0.88$). These results highlight the versatility of the DE screening approach and its potential for the time-efficient screening of libraries with minimal reagent consumption. Indeed, the whole screening process—from the transformation of the 400-variant library into *E. coli* cells to the NGS data analysis—was achieved within one week, and required only 12.5 pmol of cofactor/variant analyzed. In contrast, traditional 96-well plate screening of a library this size requires 2 nmol of cofactor/variant, which is more than a 100-fold increase in reagent consumption.



Variant	Enrichment ranking in DE screening	Activity ranking in 96-well plate screening
Sav-FR	1	2
Sav-MR	2	3
Sav-MW	3	4
Sav-LQ	4	-
Sav-FQ	5	1
Sav-MI	6	5

Figure 56. Screening of a 400-variant library in DEs and comparison with a 96-well plate assay. a) Crystal structure of the ADAse **1** · Sav S112M-K121A (PDB-ID: 6FH8)[22]. The four Sav monomers are displayed as transparent surfaces in blue, pale blue, yellow and pale yellow, respectively. The mutated positions in Sav^A (S112M^A and K121A^A, red) and the positions facing the active site in Sav^B (S112M^B and K121A^B, orange) are highlighted and labelled. The biotinylated ruthenium complex **1** is shown using a ball and stick representation and is anchored in Sav^A (blue). Dark yellow, red, blue, grey, yellow, and white spheres represent Ru, O, N, C, S, and H atoms, respectively. **b)** Coumarin FI distribution of DEs containing the library. The top 5% gate chosen for sorting is highlighted. **c)** Visual representation of the enrichment of the 400 mutants over the unsorted sample in the top 5% of the DE population. Amino acids at position K121 are represented on the x-axis, and amino acids at position S112 are represented on the y-axis. The normalized enrichment values of the respective mutants over their occurrence in the unsorted sample are displayed on a logarithmic scale. Blue: positive enrichment (values > 0), orange: negative enrichment (values < 0). **d)** Comparison of the top six mutants identified from DE screening and 96-well plate screening.

4.5.4 Conclusion

Using droplet microfluidics for the high-throughput encapsulation of live *E. coli* and readily available FACS instrumentation, we have developed a straightforward method for the rapid screening of ArMs, using a deallylase based on the biotin-streptavidin technology as a model system. We demonstrated the potential of this approach by performing an enrichment experiment

with a model library of variants, housed in the periplasm of *E. coli*, and achieving more than 68-fold enrichment of the most active variant. We further screened a library of 400 variants using the same periplasmic approach and identified five out of the six hits previously identified using a MTP screening protocol[23]. Significantly, this method requires far less time and fewer reagents than standard methods, and the screening of much larger libraries (> 500,000 variants) would not require considerably more time for the encapsulation, incubation and sorting (steps 2-4 in Figure 52b).

In summary, we have demonstrated that a single round of screening enables us to enrich moderately improved variants (i.e., ~15-fold improved activity vs WT ADAse[188]) from a medium-sized library of ArMs, compartmentalized in the periplasm of *E. coli*. This proof-of-concept lays the groundwork for future microfluidics-based directed evolution campaigns of (artificial) enzymes that afford a fluorescent (by-)product. We anticipate that the microfluidic platform presented herein will render ultra-high throughput screening of enzymatic activity accessible to virtually any research laboratory with a minimal capital investment (<30 k$) and limited microfluidic expertise.

ACKNOWLEDGEMENTS

We are grateful to all members of the FACS core facility at the Biozentrum in Basel for the great support and measurements. The Genomics Facility Basel is acknowledged for their support with NGS experiments. Dr. Basilius Sauter is kindly thanked for his support with data analysis and helpful discussions. Further, the authors thank the Swiss National Science Foundation (Grant 200020_182046), the NCCR Molecular Systems Engineering and the European Research Council (Advanced Grant DrEAM - 694424 to T. R. W. and Consolidator Grant HybCell - 681587 to P. S. D.) for their generous support.

AUTHOR CONTRIBUTION

J.V. cloned the libraries, prepared the cells for catalysis and encapsulation, performed the sorting experiments and sequencing experiments, and data visualization, and performed the *in vivo* and *in vitro* characterization on the best hits. A.S designed and fabricated the PDMS chip, performed the encapsulation experiments and flow cytometry measurements and did the data analysis. A.D.L. assisted in the CVL library design. J.V. and A.S wrote the manuscript.

SUPPORTING INFORMATION

The full supporting information can be found in Appendix C.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

4.6 REFERENCES

[2] A. Li, G. Qu, Z. Sun, M. T. Reetz, *ACS Catal.* **2019**, *9*, 7769–7778.

[3] K. Chen, F. H. Arnold, *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 5618–5622.

[4] G. Qu, A. Li, C. G. Acevedo-Rocha, Z. Sun, M. T. Reetz, *Angew. Chemie - Int. Ed.* **2020**, *59*, 13204–13231.

[6] M. T. Reetz, A. Zonta, K. Schimossek, K.-E. Jaeger, K. Liebeton, *Angew. Chemie Int. Ed. English* **1997**, *36*, 2830–2832.

[7] I. V. Pavlidis, M. S. Weiß, M. Genz, P. Spurr, S. P. Hanlon, H. Iding, U. T. Bornscheuer, *Nat. Chem.* **2016**, *8*, 1076–1082.

[8] D. S. Macdonald, X. Garrabou, C. Klaus, R. Verez, T. Mori, D. Hilvert, *J. Am. Chem. Soc* **2020**, *142*, 2020.

[13] L. Giger, S. Caner, R. Obexer, P. Kast, D. Baker, N. Ban, D. Hilvert, *Nat. Chem. Biol.* **2013**, *9*, 494–498.

[17] A. P. Green, N. J. Turner, E. O'Reilly, *Angew. Chemie Int. Ed.* **2014**, *53*, 10714–10717.

[19] S. Kille, C. G. Acevedo-Rocha, L. P. Parra, Z. G. Zhang, D. J. Opperman, M. T. Reetz, J. P. Acevedo, *ACS Synth. Biol.* **2013**, *2*, 83–92.

[23] U. Markel, K. D. Essani, V. Besirlioglu, J. Schiffels, W. R. Streit, U. Schwaneberg, *Chem. Soc. Rev.* **2020**, *49*, 233–262.

[30] F. J. Ghadessy, J. L. Ong, P. Holliger, *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 4552–4557.

[38] R. Obexer, M. Pott, C. Zeymer, A. D. Griffiths, D. Hilvert, *Protein Eng. Des. Sel.* **2016**, *29*, 355–366.

[41] A. C. Larsen, M. R. Dunn, A. Hatch, S. P. Sau, C. Youngbull, J. C. Chaput, *Nat. Commun.* **2016**, *7*, 1–9.

[70] J. F. Edd, D. Di Carlo, K. J. Humphry, S. Köster, D. Irimia, D. A. Weitz, M. Toner, *Lab Chip* **2008**, *8*, 1262–1264.

[73] M. Fischlechner, Y. Schaerli, M. F. Mohamed, S. Patil, C. Abell, F. Hollfelder, *Nat. Chem.* **2014**, *6*, 791–796.

[98] J. J. Agresti, E. Antipov, A. R. Abate, K. Ahn, A. C. Rowat, J. C. Baret, M. Marquez, A. M. Klibanov, A. D. Griffiths, D. A. Weitz, *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 4004–4009.

[99] B. Kintses, C. Hein, M. F. Mohamed, M. Fischlechner, F. Courtois, C. Lainé, F. Hollfelder, *Chem. Biol.* **2012**, *19*, 1001–1009.

[101] R. Obexer, M. Pott, C. Zeymer, A. D. Griffiths, D. Hilvert, *Protein Eng. Des. Sel.* **2016**, *29*, 355–366.

[102] R. Obexer, A. Godina, X. Garrabou, P. R. E. Mittl, D. Baker, A. D. Griffiths, D. Hilvert, *Nat. Chem.* **2017**, *9*, 50–56.

[113] K. Bernath, M. Hai, E. Mastrobattista, A. D. Griffiths, S. Magdassi, D. S. Tawfik, *Anal. Biochem.* **2004**, *325*, 151–157.

[114] E. Mastrobattista, V. Taly, E. Chanudet, P. Treacy, B. T. Kelly, A. D. Griffiths, *Chem. Biol.* **2005**, *12*, 1291–1300.

[134] S. S. Terekhov, I. V. Smirnov, A. V. Stepanova, T. V. Bobik, Y. A. Mokrushina, N. A. Ponomarenko, A. A. Belogurov, M. P. Rubtsova, O. V. Kartseva, M. O. Gomzikova, A. A. Moskovtsev, A. S. Bukatin, M. V. Dubina, E. S. Kostryukova, V. V. Babenko, M. T. Vakhitova, A. I. Manolov, M. V. Malakhova, M. A. Kornienko, A. V. Tyakht, A. A. Vanyushkina, E. N. Ilina, P. Masson, A. G. Gabibov, S. Altman, *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 2550–2555.

[135] A. Zinchenko, S. R. A. Devenish, B. Kintses, P. Y. Colin, M. Fischlechner, F. Hollfelder, *Anal. Chem.* **2014**, *86*, 2526–2533.

[143] K. K. Brower, C. Carswell-Crumpton, S. Klemm, B. Cruz, G. Kim, S. G. K. Calhoun, P. M. Fordyce, *Lab Chip* **2020**, *20*, 2062–2074.

[144] K. K. Brower, M. Khariton, P. H. Suzuki, C. Still, G. Kim, S. G. K. Calhoun, L. S. Qi, B. Wang, P. M. Fordyce, *Anal. Chem.* **2020**, *92*, 13262–13270.

[152] V. Sabatino, J. G. Rebelein, T. R. Ward, *J. Am. Chem. Soc* **2019**, *141*, 24–30.

[153] J. G. Rebelein, Y. Cotelle, B. Garabedian, T. R. Ward, *ACS Catal.* **2019**, *9*, 4173–4178.

[159] A. D. Liang, J. Serrano-Plana, R. L. Peterson, T. R. Ward, *Acc. Chem. Res.* **2019**, *52*, 585–595.

[160] F. Schwizer, Y. Okamoto, T. Heinisch, Y. Gu, M. M. Pellizzoni, V. Lebrun, R. Reuter, V. Köhler, J. C. Lewis, T. R. Ward, *Chem. Rev.* **2018**, *118*, 142–231.

[177] Y. Okamoto, V. Köhler, T. R. Ward, *J. Am. Chem. Soc.* **2016**, *138*, 5781–5784.

[179] T. Heinisch, F. Schwizer, B. Garabedian, E. Csibra, M. Jeschek, J. Vallapurackal, V. B. Pinheiro, P. Marlière, S. Panke, T. R. Ward, *Chem. Sci.* **2018**, *9*, 5383–5388.

[180] J. Serrano-Plana, C. Rumo, J. G. Rebelein, R. L. Peterson, M. Barnet, T. R. Ward, *J. Am. Chem. Soc.* **2020**, *142*, 10617–10623.

[187] M. Jeschek, R. Reuter, T. Heinisch, C. Trindler, J. Klehr, S. Panke, T. R. Ward, *Nature* **2016**, *537*, 661–665.

[188] T. Vornholt, F. Christoffel, M. M. Pellizzoni, S. Panke, T. R. Ward, M. Jeschek, *Sci. Adv.* **2021**, *7*, 4208–4230.

[190] S. Wu, Y. Zhou, J. G. Rebelein, M. Kuhn, H. Mallin, J. Zhao, N. V. Igareta, T. R. Ward, *J. Am. Chem. Soc.* **2019**, *141*, 15869–15878.

[191] F. Christoffel, N. V Igareta, M. M. Pellizzoni, L. Tiessler-Sala, B. Lozhkin, D C. Spiess, A. Lledose, J.-D. Maréchal, R. L. Peterson, T. R. Ward, Nature Catlysis, accepted.

[192] K. Oohora, A. Onoda, T. Hayashi, *Acc Chem Res* **2019**, *52*, 945–954.

[193] G. Roelfes, *Acc. Chem. Res.* **2019**, *52*, 545–556.

[195] J. G. Rebelein, Y. Cotelle, B. Garabedian, T. R. Ward, *ACS Catal.* **2019**, *9*, 4173–4178.

[210] C. Streu, E. Meggers, *Angew. Chemie - Int. Ed.* **2006**, *45*, 5645–5648.

[211] T. Völker, F. Dempwolff, P. L. Graumann, E. Meggers, *Angew. Chem. Int. Ed. Engl.* **2014**, *53*, 10536–10540.

[213] T. Völker, E. Meggers, *ChemBioChem* **2017**, *18*, 1083–1086.

[223] T. Heinisch, F. Schwizer, B. Garabedian, E. Csibra, M. Jeschek, J. Vallapurackal, V. B. Pinheiro, P. Marlière, S. Panke, T. R. Ward, *Chem. Sci.* **2018**, *9*, 5383–5388.

[293] P. A. Romero, F. H. Arnold, *Nat. Rev. Mol. Cell Biol.* **2009**, *10*, 866–876.

[294] N. Tokuriki, D. S. Tawfik, *Science.* **2009**, *324*, 203–207.

[296] P. C. Phillips, *Nat. Rev. Genet.* **2008**, *9*, 855–867.

[297] J. M. Smith, *Nature* **1970**, *225*, 563–564.

[298] S. A. Kauffman, E. D. Weinberger, *J. Theor. Biol.* **1989**, *141*, 211–245.

[299] C. Yan, F. Parmeggiani, E. A. Jones, E. Claude, S. A. Hussain, N. J. Turner, S. L. Flitsch, P. E. Barran, *J. Am. Chem. Soc* **2017**, *139*, DOI 10.1021/jacs.6b12165.

[300] J. W. Janiesch, M. Weiss, G. Kannenberg, J. Hannabuss, T. Surrey, I. Platzman, J. P. Spatz, *Anal. Chem.* **2015**, *87*, 2063–2067.

[301] G. Woronoff, A. El Harrak, E. Mayot, O. Schicke, O. J. Miller, P. Soumillion, A. D. Griffiths, M. Ryckelynck, *Anal. Chem.* **2011**, *83*, 2852–2857.

[302] S. Deshpande, Y. Caspi, A. E. Meijering, C. Dekker, *Nat. Commun.* **2016**, *7*, 10447.

[303] A. Stucki, P. Jusková, N. Nuti, S. Schmitt, P. S. Dittrich, *Small Methods* **2021**, *5*, 2100331.

[304] "The Technology Behind TOPO Cloning | Thermo Fisher Scientific - CH," can be found under https://www.thermofisher.com/ch/en/home/life-science/cloning/topo/topo-resources/the-technology-behind-topo-cloning.html, **n.d.**

[305] S. Fanning, S. Proos, K. Jordan, S. Srikumar, *Front. Microbiol.* **2017**, *8*, 1829.

[306] M. Fedurco, A. Romieu, S. Williams, I. Lawrence, G. Turcatti, *Nucleic Acids Res.* **2006**, *34*, e22–e22.

[307] C. Adessi, G. Matton, G. Ayala, G. Turcatti, J. J. Mermod, P. Mayer, E. Kawashima, *Nucleic Acids Res.* **2000**, *28*, 87.

[308] G. Turcatti, A. Romieu, M. Fedurco, A. P. Tairi, *Nucleic Acids Res.* **2008**, *36*, 25.

[309] B. Canard, R. S. Sarfati, *Gene* **1994**, *148*, 1–6.

[310] J. Hagmann, Characterizing Short-Term Evolution of DNA Methylation in A. Thaliana Using next-Generation Sequencing, **2017**.

[311] L. R. Runtuwene, J. S. B. Tuda, A. E. Mongan, W. Makalowski, M. C. Frith, M. Imwong, S. Srisutham, L. A. Nguyen Thi, N. N. Tuan, Y. Eshita, R. Maeda, J. Yamagishi, Y. Suzuki, *Sci. Rep.* **2018**, *8*, 1–13.

[312] S. L. Castro-Wallace, C. Y. Chiu, K. K. John, S. E. Stahl, K. H. Rubins, A. B. R. McIntyre, J. P. Dworkin, M. L. Lupisella, D. J. Smith, D. J. Botkin, T. A. Stephenson, S. Juul, D. J. Turner, F. Izquierdo, S. Federman, D. Stryke, S. Somasekar, N. Alexander, G. Yu, C. E. Mason, A. S. Burton, *Sci. Rep.* **2017**, *7*, 1–12.

[313] R. A. Bull, T. N. Adikari, J. M. Ferguson, J. M. Hammond, I. Stevanovski, A. G. Beukers, Z. Naing, M. Yeang, A. Verich, H. Gamaarachchi, K. W. Kim, F. Luciani, S. Stelzer-Braid, J. S. Eden, W. D. Rawlinson, S. J. van Hal, I. W. Deveson, *Nat. Commun.* **2020**, *11*, 1–8.

[314] M. Akeson, D. Branton, G. Church, D. W. Deamer, *United States (12) Patent Application Publication (10) Pub*, **2012**.

[315] S. Howorka, S. Cheley, H. Bayley, *Nat. Biotechnol.* **2001**, *19*, 636–639.

[316] L. Xu, M. Seki, *J. Hum. Genet.* **2020**, *65*, 25–33.

[317] T. Vornholt, F. Christoffel, M. Pellizzoni, S. Panke, T. Ward, M. Jeschek, *bioRxiv* **2020**, 2020.07.15.204206.

[318] T. Vornholt, F. Christoffel, M. M. Pellizzoni, S. Panke, T. R. Ward, M. Jeschek, *Sci. Adv.* **2021**, *7*, 4208–4230.

[319] Z. Liu, F. H. Arnold, *Curr. Opin. Biotechnol.* **2021**, *69*, 43–51.

[320] S. N. Natoli, J. F. Hartwig, *Acc. Chem. Res.* **2019**, *52*, 326–335.

[321] J. C. Lewis, *Acc. Chem. Res.* **2019**, *52*, 576–584.

[322] A. Lombardi, F. Pirro, O. Maglio, M. Chino, W. F. DeGrado, *Acc. Chem. Res.* **2019**, *52,* 1148–1159.

[323] C. S. Mocny, V. L. Pecoraro, *Acc. Chem. Res.* **2015**, *48*, 2388–2396.

[324] A. Stucki, J. Vallapurackal, T. R. Ward, P. S. Dittrich, *Angew. Chem. Int. Ed.* **2021**, *60*, 24368–24387.

[325] J. M. Holstein, C. Gylstorff, F. Hollfelder, *ACS Synth. Biol.* **2021**, *10*, 252–257.

[325] D. Vallejo, A. Nikoomanzar, B. M. Paegel, J. C. Chaput, *ACS Synth. Biol.* **2019**, *8*, 1430–1440.

[326] M. Jeschek, M. O. Bahls, V. Schneider, P. Marlière, T. R. Ward, S. Panke, *Metab. Eng*. **2017**, *40*, 33–40.

SUMMARY & OUTLOOK

Directed evolution has become a common tool for the optimization of virtually any given protein or DNA. This doctoral thesis summarizes our efforts to take directed evolution of ArMs a step further by the means of droplet microfluidics, *in vivo* screening, FACS and high-throughput sequencing.

An ultrahigh-throughput screening assay to screen and analyze > 0.5 Mio variants within one week was developed. By exploiting the high-throughput of droplet microfluidics to encapsulate live *E. coli* cells, and coupling it to FACS instrumentation, we developed a straightforward method for the genetic engineering of ArMs. Screening of a quadruple mutant library, resulted in three highly active variants displaying significantly-improved catalytic activities over their single and double mutation variants. In conclusion, this work represents a powerful method for a single-round fast engineering and remodeling of an active site. We predict that such a screening platform will facilitate and speed-up the genetic optimization of ArMs *in vivo*.

Engineering enzymatic activity is particularly challenging since even the tiniest changes in a protein's sequence can have big effects on its function. Thus, predicting how a given mutation might change a behavior remains a challenge. Nature found a way to optimize certain reactions over billions of years, but a major throwback in the process is, that we don't know why or how we ended up with the present enzymes. Glancing at the vast possibilities –which is more than the number of atoms in our universe– one can only dream of what might be still possible to achieve if only we could understand proteins better.

I strongly believe that screenings of the type and size we report here, are just a first step in this direction. Screening and analysis of function improving hits as well as function-depleting non-hits can serve as valuable training sets for future machine learning algorithms; ultimately enabling us to see more than just mere sequence and activity, but to understand how sequence affects function.

APPENDIX A

A.1 SUPPLEMENTARY FIGURES AND TABLES

A.1.1 ARTIFICIAL DEALLYLASE

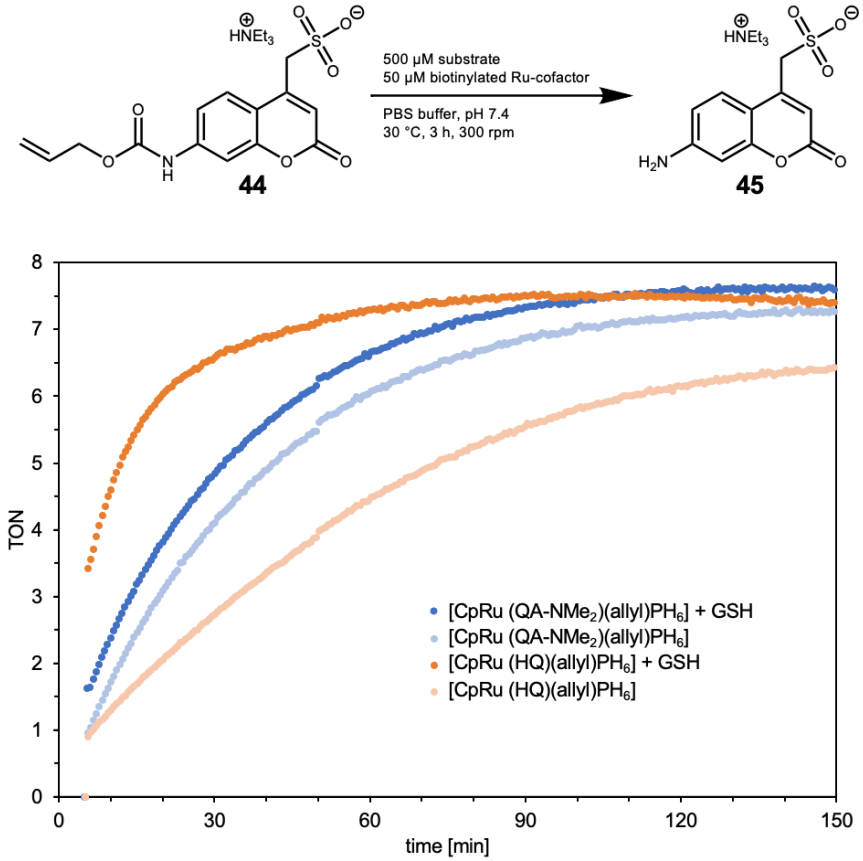
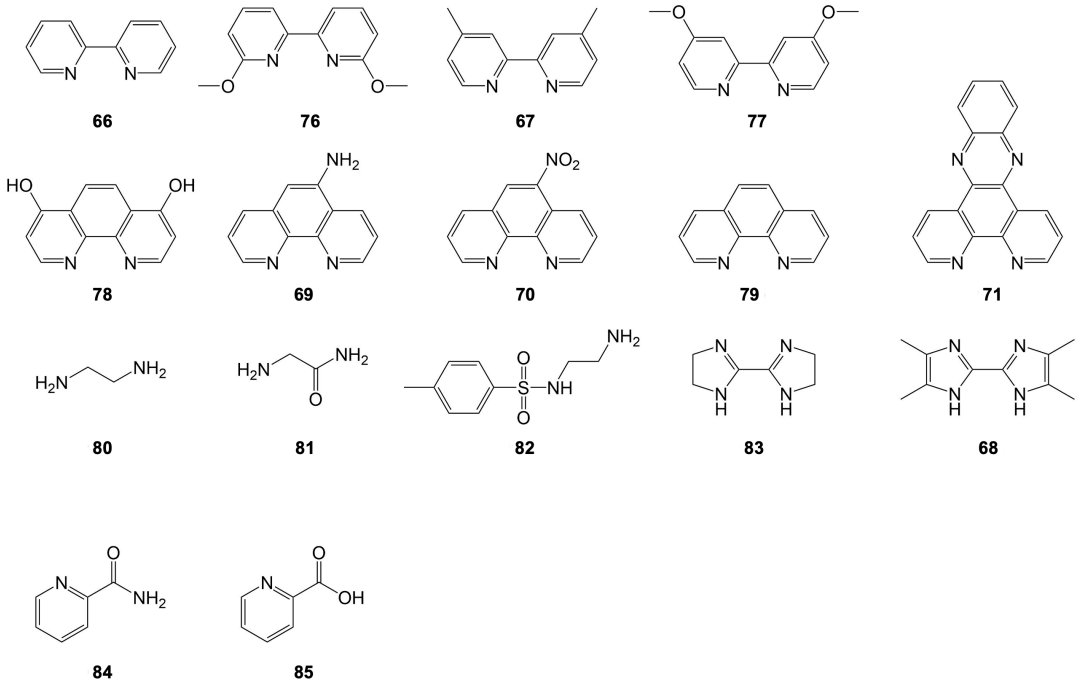


Figure S1. Time point measurements for the ruthenium catalyzed deprotection of allyl-carbamate protected coumarin **44** using [CpRu(QA)(Allyl)]PF_6 **38** and [CpRu(HQ)(Allyl)]PF_6 **39**. Reaction conditions: **a)** substrate **44** (500 µM), ruthenium catalyst in water and DMF (0.5%) 30 °C, air. **b)** substrate **44** (500 µM), ruthenium catalyst in water and DMF (0.5%), GSH (50 mM), 30 °C, air.

A.1.2 ARTIFICIAL CARROLLASE



Scheme S1. Ligands based on bpy and phenanthroline with *N*,*N*-diamine motifs.

Table S1. Summary of ligand screening using bpy and phenanthroline derivatives.

Entry	ligand	TON		
		10 mol%	5 mol%	1 mol%
1	**66**	0.99	0.96	1.06
2	**76**	0.17	0.17	-
3	**67**	1.56	1.38	1.94
4	**77**	0.74	0.67	-
5	**78**	0.73	0.59	-
6	**69**	2.69	2.76	5.54
7	**70**	3.54	4.33	8.06
8	**79**	2.00	1.59	6.47
9	**71**	2.41	3.75	-
10	**80**	-	-	-
11	**81**	-	-	-
12	**82**	-	-	-
13	**83**	0.18	0.16	1.52
14	**68**	1.65	2.11	-
15	**84**	0.48	0.46	-
16	**85**	0.35	0.28	-

Reaction conditions: substrate **52** (2 mM), CpRu-**XX** catalyst, PBS + DMF (10%), 37 °C, 17 h.

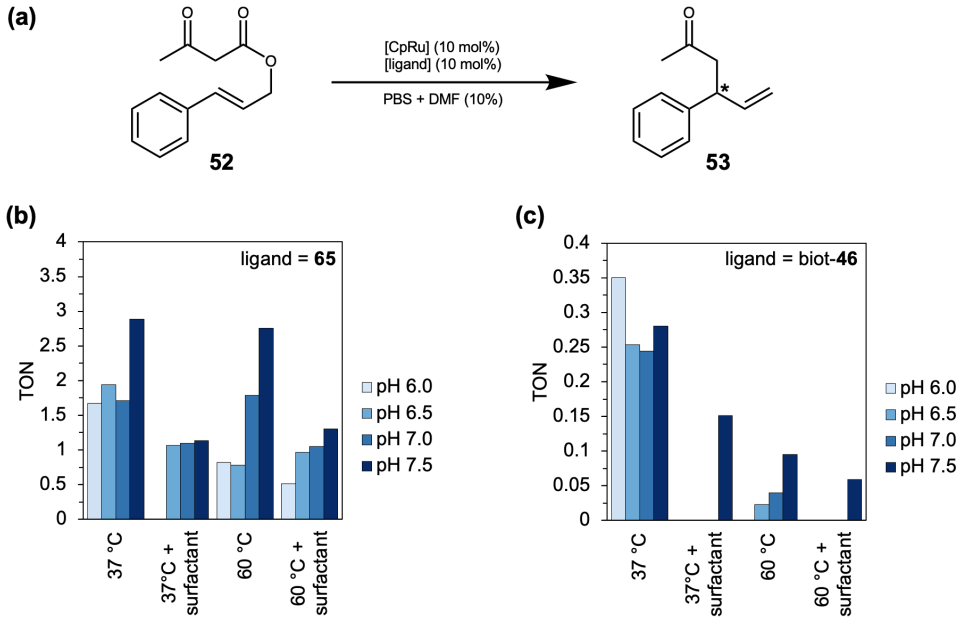


Figure S2. Screening of different pH and temperatures, as well as the addition of the surfactant TPGS-750-M. Reaction conditions: **b)** substrate **52** (2 mM), ruthenium catalyst (200 µM), ligand **65** (200 µM), surfactant TPGS-750-M, water + DMF (10%), pH 6.0-7.5, 37-60 °C, 17 h. **c)** substrate **52** (2 mM), ruthenium catalyst (200 µM), ligand (200 µM), Sav (400 µM), surfactant TPGS-750-M, water + DMF (10%), pH 6.0-7.5, 37-60 °C, 17 h.

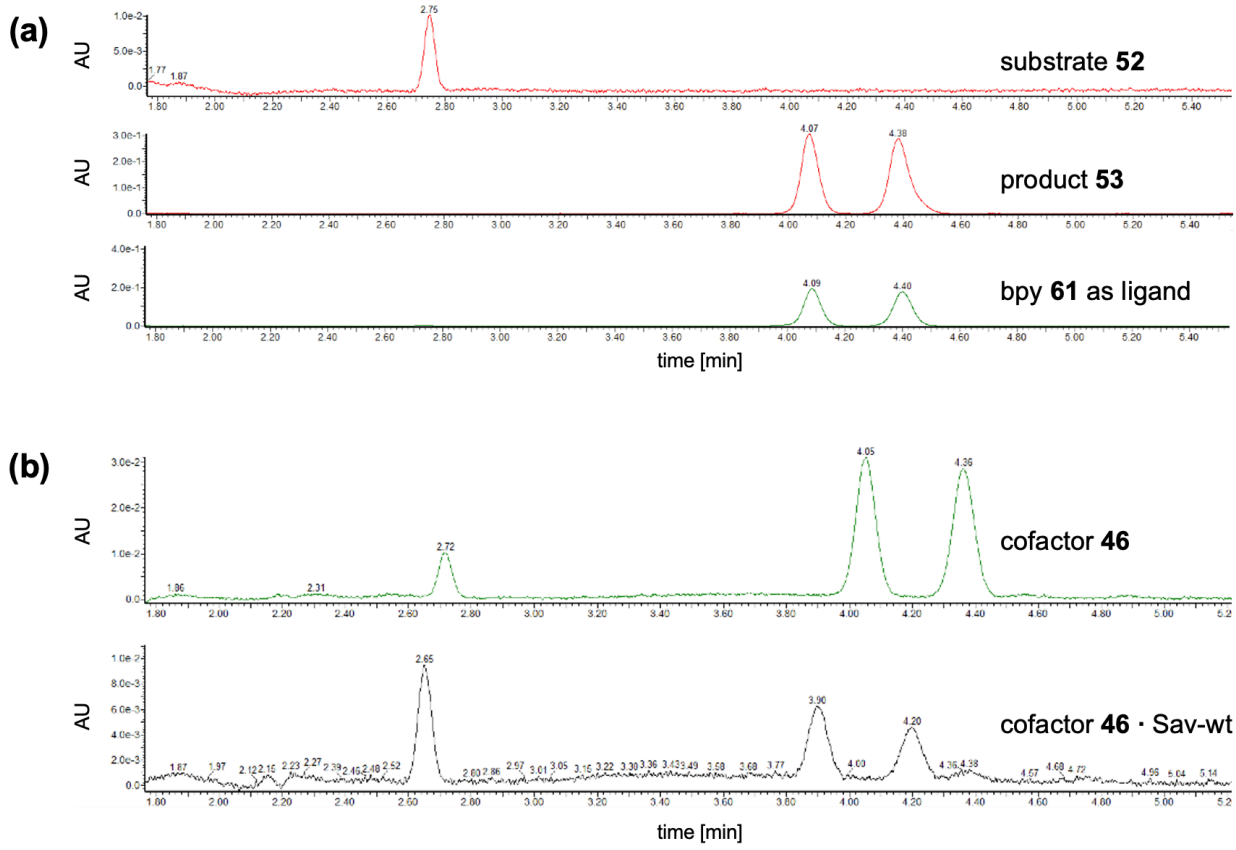


Figure S3. Analysis of the Carroll rearrangement via UPC². a) pure substrate, pure product and a control reaction using bpy **61** as a ligand under "standard Lacour conditions". Under the standard conditions, full conversion without any enantioselectivity (~1.66 *ee*) is observed b) Catalysis using only cofactor **46** (top) and cofactor **46** · Sav-wt (bottom) in water + DMF (10%). The conversions are significantly lower. However, in the presence of protein, a slight enantioselectivity (~16 *ee*) can be seen. But because of the low conversion, this has to be interpreted with caution.

Table S2. *In vitro* screening of a small Sav library.

Entry	Complex/ArM	catalyst loading [mol%]	cofactor 46 TON	cofactor 47 TON
1	wt	10	0.09	0.08
2	Sav-YR	10	0.04	0.11
3	S112A	10	0.13	0.06
4	S112C	10	0.07	0.04
5	S112E	10	-	0.03
6	S112H	10	0.14	0.09
7	S112K	10	-	0.04
8	S112M	10	0.14	0.09
9	S112N	10	0.19	0.06
10	S112R	10	0.12	0.04
11	S112T	10	0.07	0.05
12	S112W	10	-	0.06
13	S112Y	10	0.10	0.09
14	K121A	10	0.21	0.24
15	K121C	10	0.04	0.08
16	K121F	10	0.09	0.18
17	K121H	10	-	0.06
18	K121I	10	0.29	0.08
19	K121M	10	0.63	0.13
20	K121P	10	0.22	0.16
21	K121Q	10	0.21	0.05
22	K121R	10	0.03	0.09
23	K121S	10	0.23	0.15
24	K121T	10	0.23	0.13
25	K121V	10	0.31	0.24
26	K121W	10	0.28	0.17
27	K121Y	10	0.17	0.18
28	L124C	10	0.04	0.05
29	L124F	10	-	0.04
30	L124G	10	0.17	0.23
31	L124H	10	-	0.03
32	L124K	10	-	0.04
33	L124M	10	0.08	0.06
34	L124N	10	-	0.05
35	L124V	10	0.05	0.05
36	L124Y	10	-	0.02

Reaction conditions: substrate **52** (2 mM), CpRu-biot-**46** catalyst (200 µM), Sav (400 µM), water + DMF (10%), 37 °C, 17 h.

A.2 Experimental Part

A.2.1 Materials and Methods

All commercially available chemicals were purchased from Sigma-Aldrich, ABCR, TCI Europe, Acros Organics, Alfa Aesar, Fluka or Merck and were used without further purification. Deuterated solvents were purchased from Cambridge Isotope Laboratories Inc. Solvents for UPLC were purchased from Romil. Water was purified with a Milli-Q-system (Millipore).

All chemical experiments were performed in oven-dried glassware. If not otherwise stated all reactions were set under an inert atmosphere with nitrogen using a Schlenk line. Reactions were monitored by TLC (Merck silica gel 60 F254 plates) and visualized by UV-light (254 or 360 nm) or potassium permanganate (KMnO4). Purification of the products was done by column chromatography using SiliCycle silica gel 60 (230-400 Mesh). NMR spectra were measured on Varian Gemini Bruker DPX 400 MHz at room temperature and evaluated with MestReNova. Chemical shifts (δ) are reported in parts per million (ppm) and referenced to the residual solvent peaks. Scalar coupling (J) is reported in Hertz (Hz).

For all biological experiments the equipment was sterilised (121 °C, 20 min). Sterilisation was done with a Tuttnauer 3850 EL autoclave. Ultrahigh-performance liquid chromatography (UPLC) was done on an Aquity UPLC H-Class Bio from Waters using a BEH C18 column (1.7 µm, 2.1 x 50 mm). Supercritical CO_2 based reverse-phase liquid chromatography was done on an Aquity UPC^2 from Waters.

A.2.2 Synthesis

[CpRu(QA-NMe$_2$)(Allyl)]PF$_6$ ***38***

The synthesis of [CpRu(QA-NMe$_2$)(Allyl)]PF_6 (**38**) was performed as reported by Völker *et al.* and was kindly provided by Dr. Fabian Schwizer.

[CpRu(HQ)(Allyl)]PF$_6$ ***39***

The synthesis of [CpRu(HQ)(Allyl)]PF_6 (**39**) was performed as reported by Völker *et al.*

^{1}H-NMR (500 MHz, DMSO-d_6) δ 9.65 (d, J = 8.8 Hz, 1H), 9.09 (d, J = 4.8 Hz, 1H), 8.23 (d, J = 8.8 Hz, 1H), 7.77 (d, J = 11.1 Hz, 1H), 6.95 (d, J = 8.6 Hz, 1H), 6.38 (s, 5H), 4.80 (m, 1H), 4.72 (d, J = 11.4 Hz, 1H), 4.52 (d, J = 11.6 Hz, 1H), 4.46 (dd, J = 5.8, 2.6 Hz, 1H), 4.26 (dd, J = 2.5 Hz, 1H), 3.89 (s, 3H) ppm.

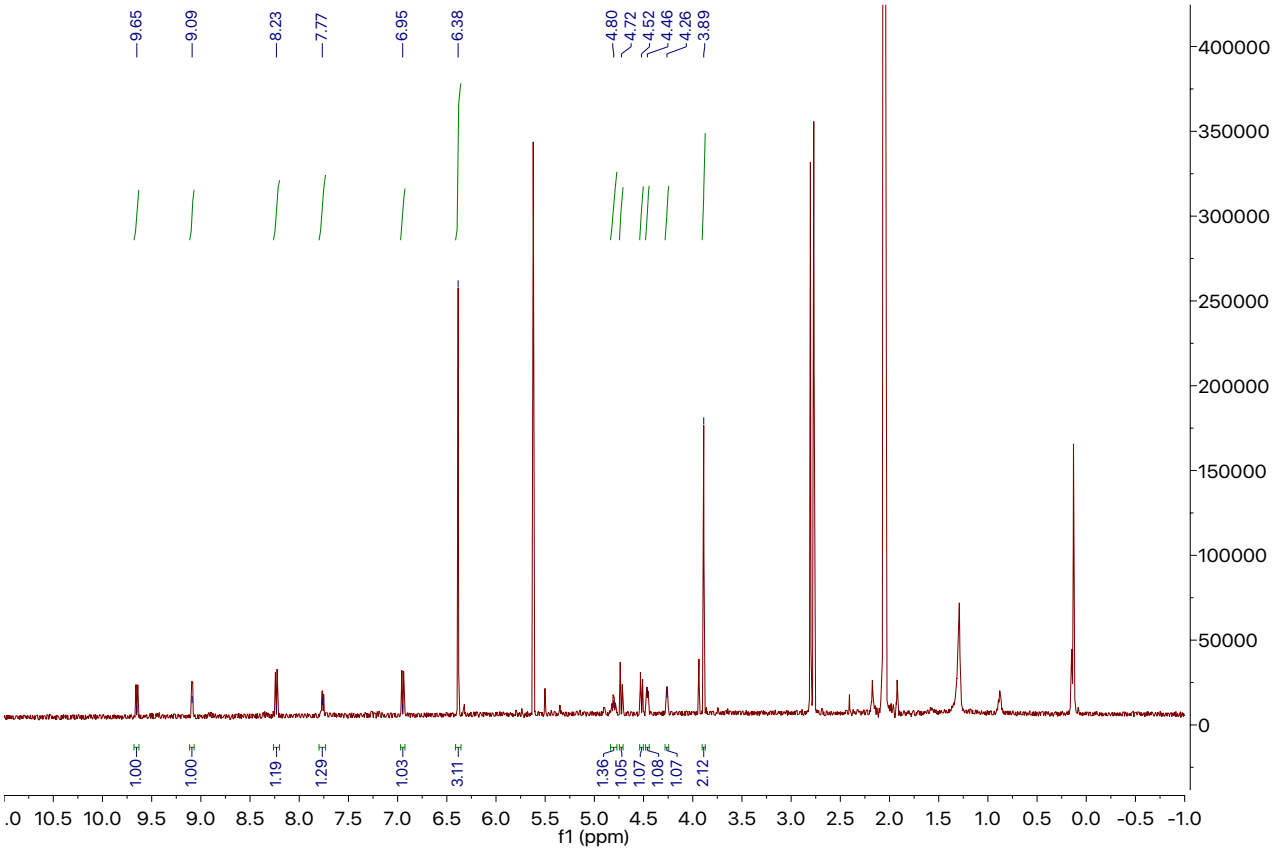


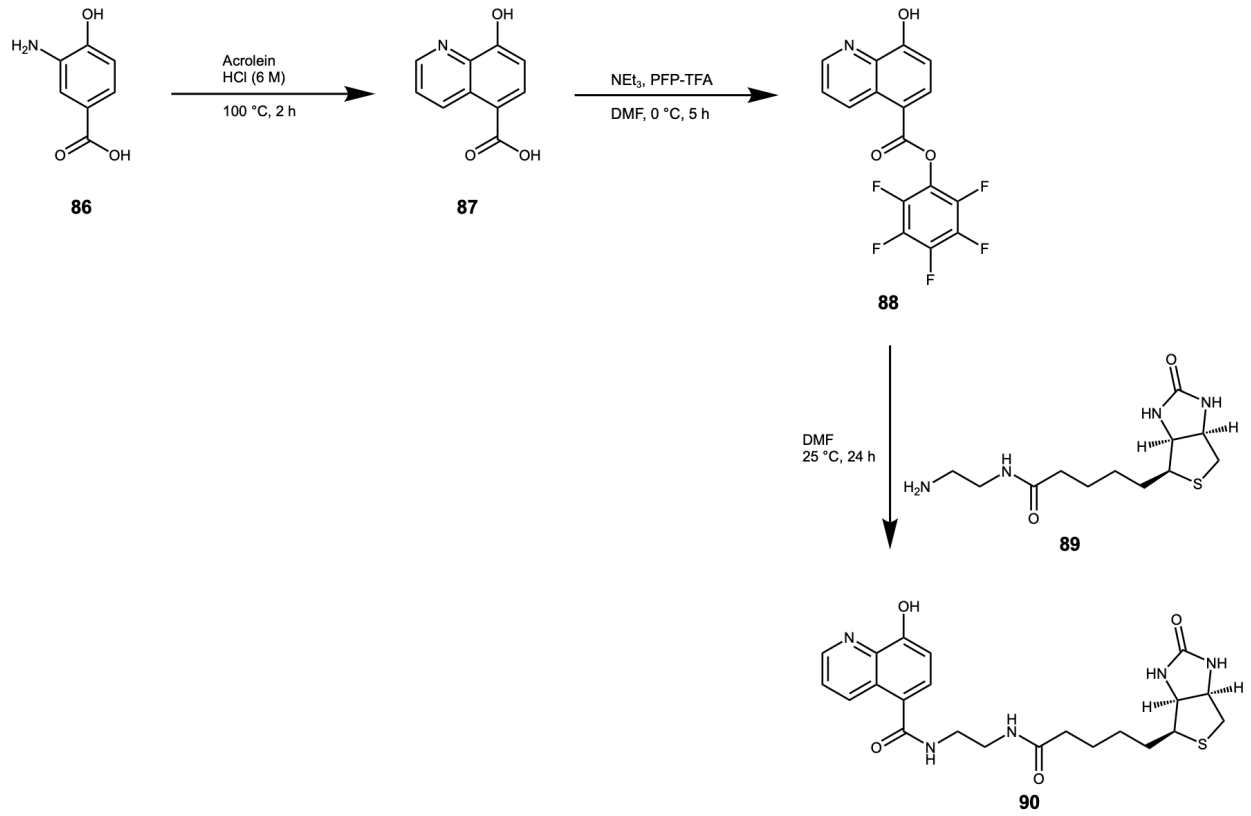
Figure S4. ^{1}H-NMR of compound **39**.

*LIGAND FOR COFACTOR **46***

The synthesis of the ligand was performed as reported by Heinisch and Schwizer *et al.*

^{1}H-NMR (500 MHz, DMSO-d_6) δ 8.12 (d, J = 8.5 Hz, 1H), 8.06 (d, J = 8.5 Hz, 1H), 7.71 (dd, J = 8.2, 6.9 Hz, 1H), 7.64 (s, 1H), 7.57 (dd, J = 8.2, 6.9, Hz, 1H), 4.49 (dd, J = 7.7, 4.6 Hz, 1H), 4.32 (dd, J = 7.7, 5.0 Hz, 1H), 3.91 (t, J = 5.5 Hz, 2H), 3.86 (t, J = 4.9 Hz), 3.39 – 3.23 (m, 4H), 3.27 – 3.18 (m, 1H), 2.95 (d, J = 5.0 Hz, 1H), 2.71 (d, J = 7.9 Hz, 1H), 2.55 (t, J = 7.4 Hz, 2H), 1.83 – 1.57 (m, 4H), 1.55 – 1.45 (m, 2H).

Figure S5. [1]H-NMR of the ligand for cofactor **46**.



Scheme S2. Synthesis overview of the biotinylated ligand **85** for the preparation of cofactor **47.**

*COMPOUND **87***

Acrolein (1.5 eq.) was added dropwise over 30 min to a solution of 3-amino-4-hydroxybenzoic acid (**86**, 1.0 eq.) in 6 N HCl. The reaction mixture was refluxed at 100 °C overnight in a 50 mL round bottom flask fitted with a jacketed water condenser. Upon completion of the reaction determined by TLC, the reaction mixture was allowed to cool to room temperature and the pH was adjusted to pH 9 with aqueous ammonia. The mixture was then filtered. The filtrate was acidified to pH 4-5 with 10% aqueous acetic acid. The resulting precipitate was obtained by filtration, washed with water (10 mL) and dried in vacuo to yield the desired product **87** as a brown powder.

^{1}H NMR (500 MHz, DMSO-d_6) δ 9.58 (d, J = 9.6 Hz, 1H), 8.86 (d, J = 4.8 Hz, 1H), 8.17 (d, J = 8.6 Hz, 1H), 7.63 (dd, J = 9.2, 4.4 Hz, 1H), 7.08 (d, J = 8.6 Hz, 2H).

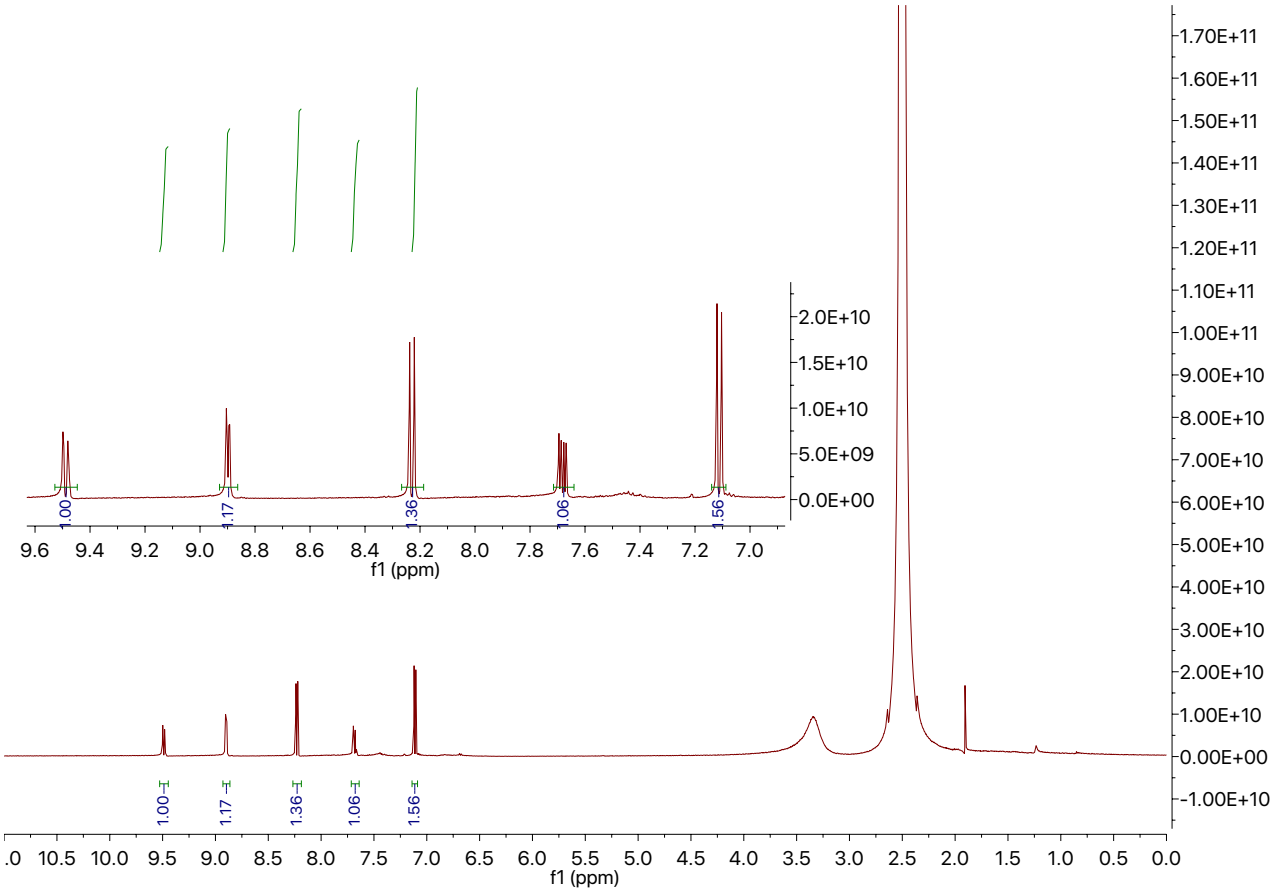


Figure S6. ^{1}H-NMR of compound **82**.

*COMPOUND **90***

To a dispersion of **87** (1.0 eq.) in DMF (3 ml) triethylamine (1.8 eq.) was added at 0 °C. Pentafluorophenyltrifluoroacetate (1.0 eq) was slowly added. The reaction mixture was allowed to warm up to room temperature and was further stirred for 5 h. The completion of the reaction was determined by TLC. The reaction mixture was used for the next step without further purification.

86 (1.0 eq.) was dissolved in DMF (5 ml) and trimethylamine (1.8 eq.) was added at 0 °C. The reaction was stirred at room temperature for 1-2 h. Afterwards, the crude reaction mixture containing the compound **88** was added slowly. The reaction mixture was allowed to warm up to room temperature and was further stirred for 24 h. DMF was evaporated and purification was done by preparative HPLC (slow gradient H_2O/MeCN with 0.1% formic acid) to yield a light brown solid.
^{1}H-NMR (400 MHz, DMSO-d_6) δ 10.24 (s, 1H), 8.87 (s, 1H), 8.42 (s, 1H), 7.88 (s, 1H), 7.66 (s, 1H), 7.58 (s, 1H), 7.11 (s, 1H), 4.24 (s, 1H), 4.07 (s, 1H), 2.97 (s, 1H), 2.72 (s, 1H), 2.52 (s, 1H), 2.08 (s, 1H), 1.61 – 1.38 (m, 4H), 1.34 – 1.21 (m, 3H).

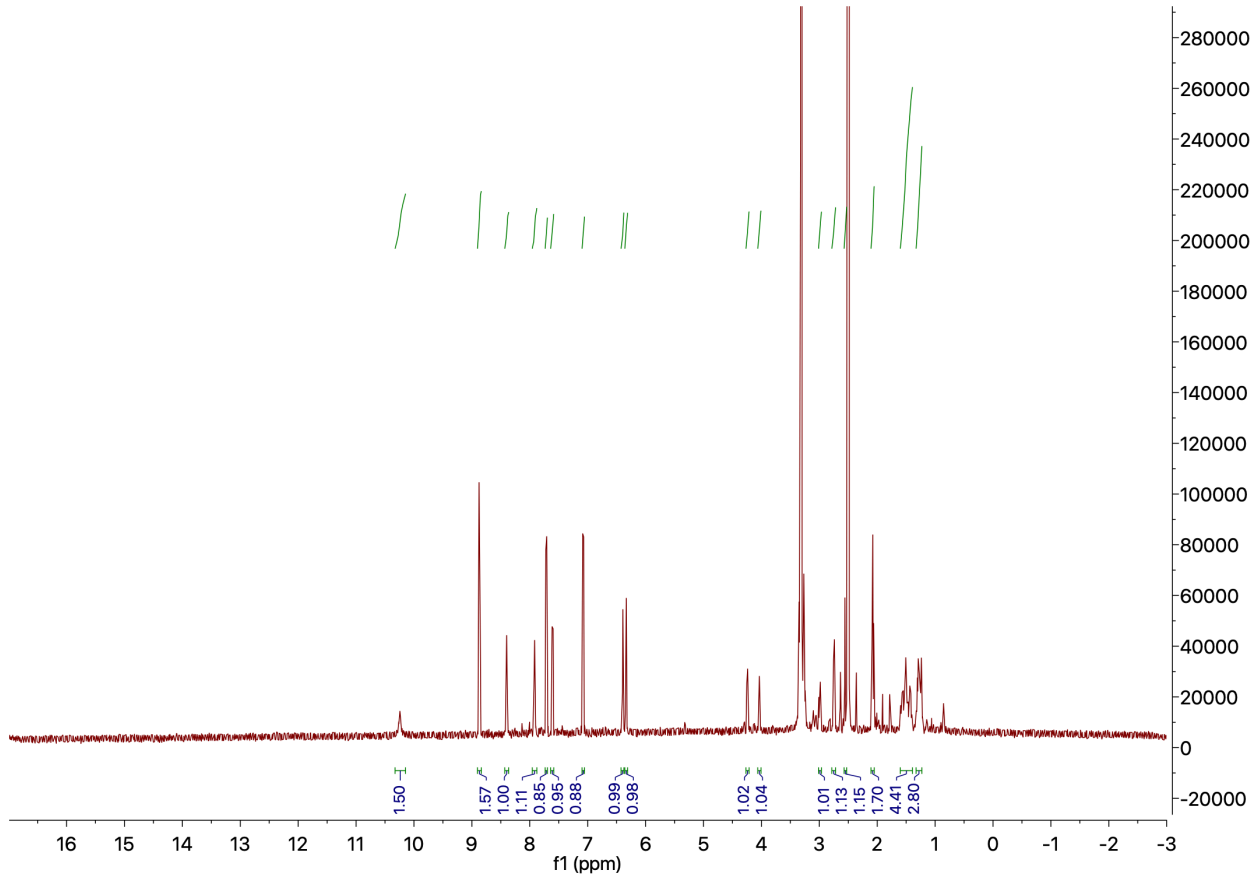


Figure S7. ^{1}H-NMR of compound **85**.

*COMPOUND **45***

7-aminocoumarin-4-methansulfonic acid (400 mg, 0.80 mmol, 1.0 eq.) was dispersed in TEAB-buffer (1M, 8.0ml, 8.00 mmol, 10 eq.) and acidified with acetic acid (100%, ~3ml) to pH 4 and then filtrated. The filtrate was then purified by preparative reverse phase HPLC (both MeCN and H2O with 50 µM formic acid and 50 µM NEt_3 as eluent).

^{1}H-NMR (400 MHz, Deuterium Oxide) δ 8.85 (s, 1H), 7.55 (d, J = 8.5 Hz, 1H), 6.50 (d, J = 8.4, 1H), 6.38 (s, 1H), 6.10 (s, 2H), 5.91 (s, 1H), 3.79 (s, 2H), 3.09 (s, 6H), 1.20 (t, J = 7.8 Hz, 9H).

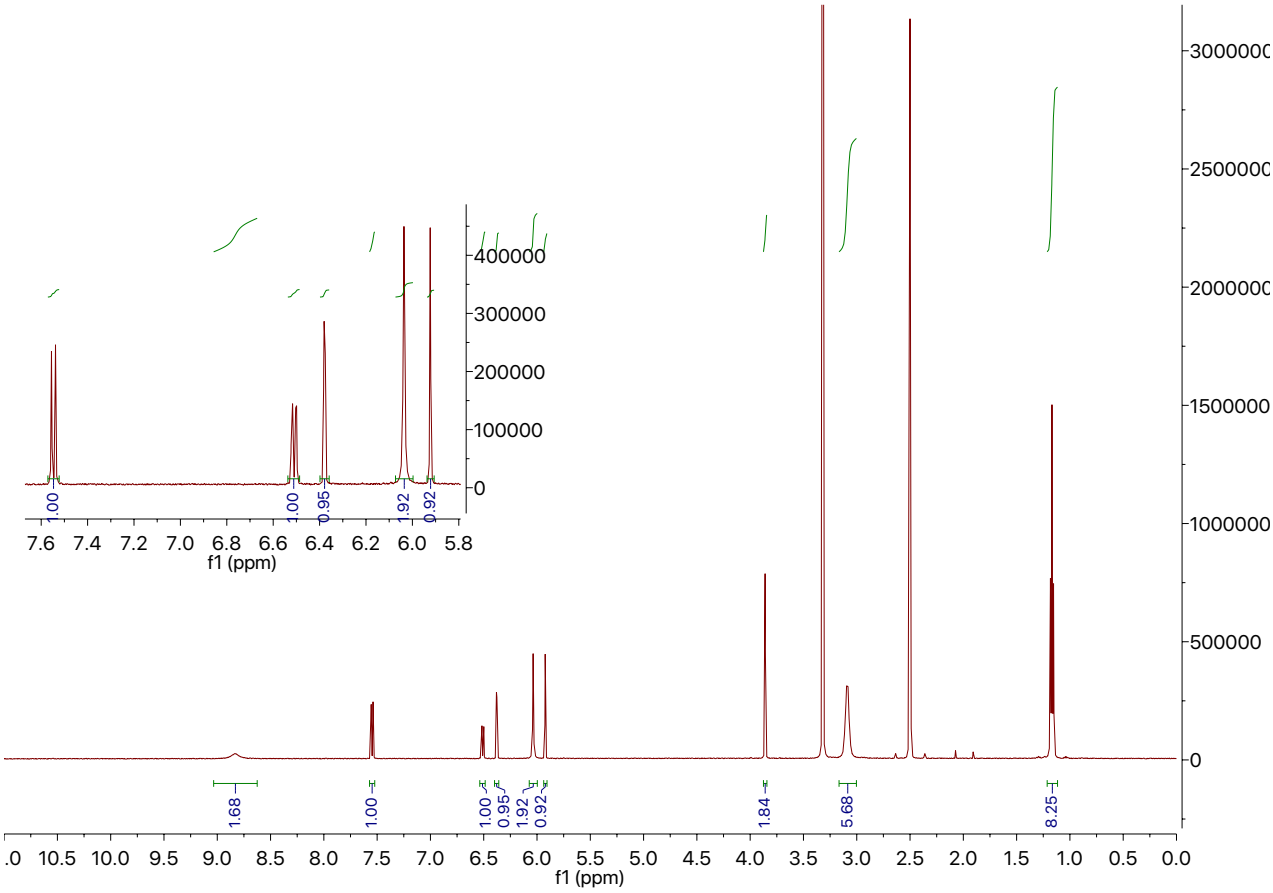


Figure S8. [1]H-NMR of compound **45**.

*COMPOUND **44***

7-aminocoumarin-4-methansulfonic acid (400 mg, 0.80 mmol, 1.0 eq.) was dispersed in TEAB-buffer (1M, 8.0ml, 8.00 mmol, 10 eq.) was added, yielding a greenish mixture. Allyl chloroformate (432 µl, 4.00 mmol, 5.0 eq.) was dropwise added at 0 °C. The mixture was then stirred for 1.5 h at 0 °C and then for 3 h at RT, yielding a yellow-brown solution. This solution was acidified with acetic acid (100%, ~3ml) to pH 4 and then filtrated. The filtrate was then purified by preparative reverse phase HPLC (both MeCN and H2O with 50 µM formic acid and 50 µM NEt3 as eluent).

[1]H-NMR (400 MHz, Deuterium Oxide) δ 7.73 (d, J = 8.9 Hz, 1H), 7.41 (d, J = 2.2 Hz, 1H), 7.19 (dd, J = 8.8, 2.2 Hz, 1H), 6.41 (s, 1H), 6.01 (ddt, J = 17.3, 10.7, 5.4 Hz, 1H), 5.39 (dq, J = 17.3, 1.6 Hz, 1H), 5.30 (dq, J = 10.6, 1.4 Hz, 1H), 4.67 (dt, J = 5.4, 1.5 Hz, 2H), 4.39 (s, 2H), 3.18 (q, J = 7.3 Hz, 6H), 1.25 (t, J = 7.3 Hz, 9H).

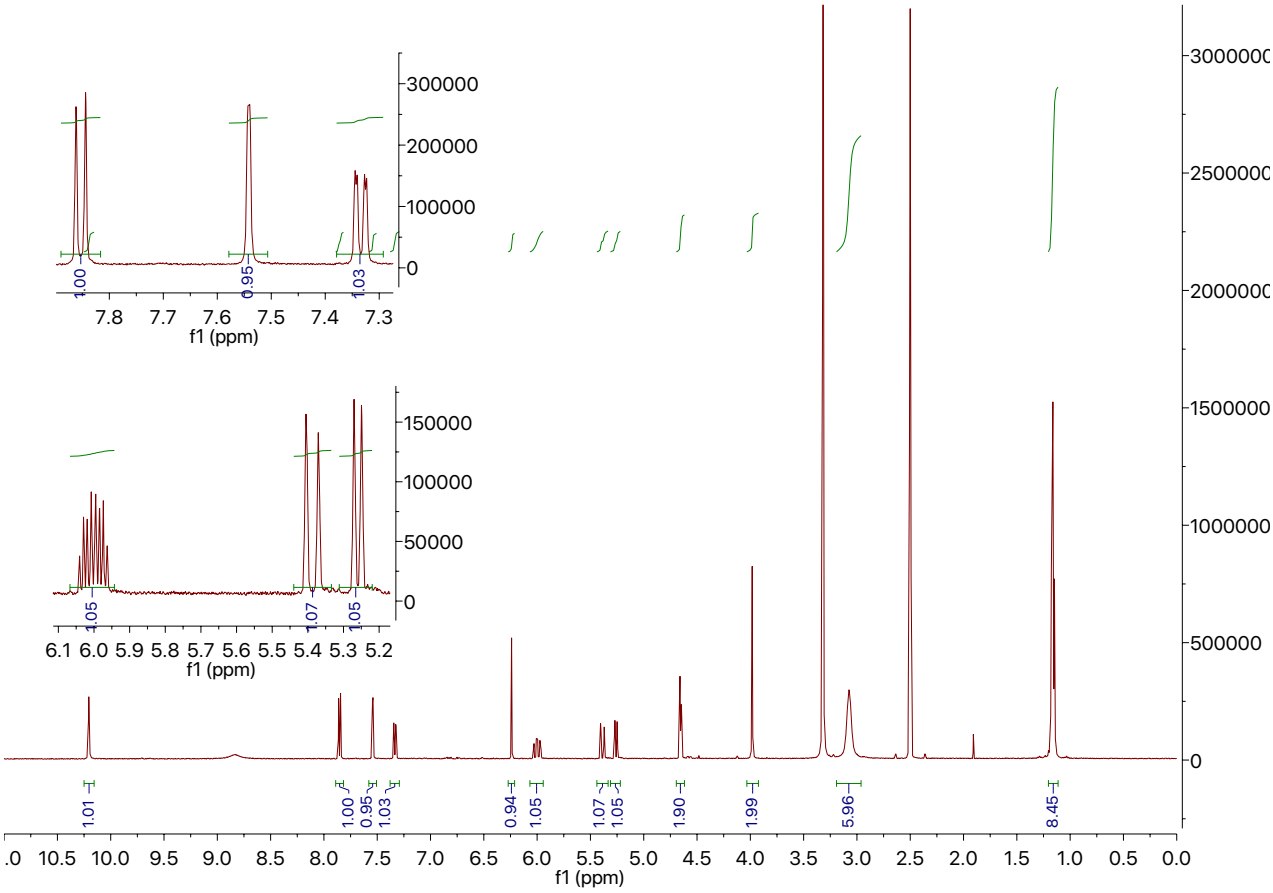


Figure S9. [1]H-NMR of compound **44**.

COMPOUND ***52***

A microwave tube was charged with $MnSO_4$, cinnamyl alcohol and ethyl acetoacetate. Toluene (10 mL) was added and the mixture was reacted in a microwave reactor for 1 h at 150 °C. Following the solvent was evaporated, and dried on the Schlenk line to ensure that most of the toluene is evaporated. The residue was then extracted with DCM and the solvent again evaporated. The crude product was obtained as a yellow oil. Final purification was done by gradient column chromatography (start: 100% cyclohexane, then 20:1 cyclohexane/EtOAc until the first fraction is out, then slowly increase to 5:1 cyclohexane/EtOAc to collect the desired product.

^{1}H-NMR (400 MHz, DMSO-d_6) δ 7.38 (d, J = 75.2 Hz, 2H), 7.29 (t, J = 5.4 Hz, 2 Hz), 7.25 (d, J = 4.9 Hz, 1H), 6.66 (d, J = 11.3), 6.27 (dd, J = 8.9, 4.7 Hz, 1H), 4.79 (d, J = 4.8 Hz, 2H), 3.49 (s, 2H), 2.35 (s, 3H).

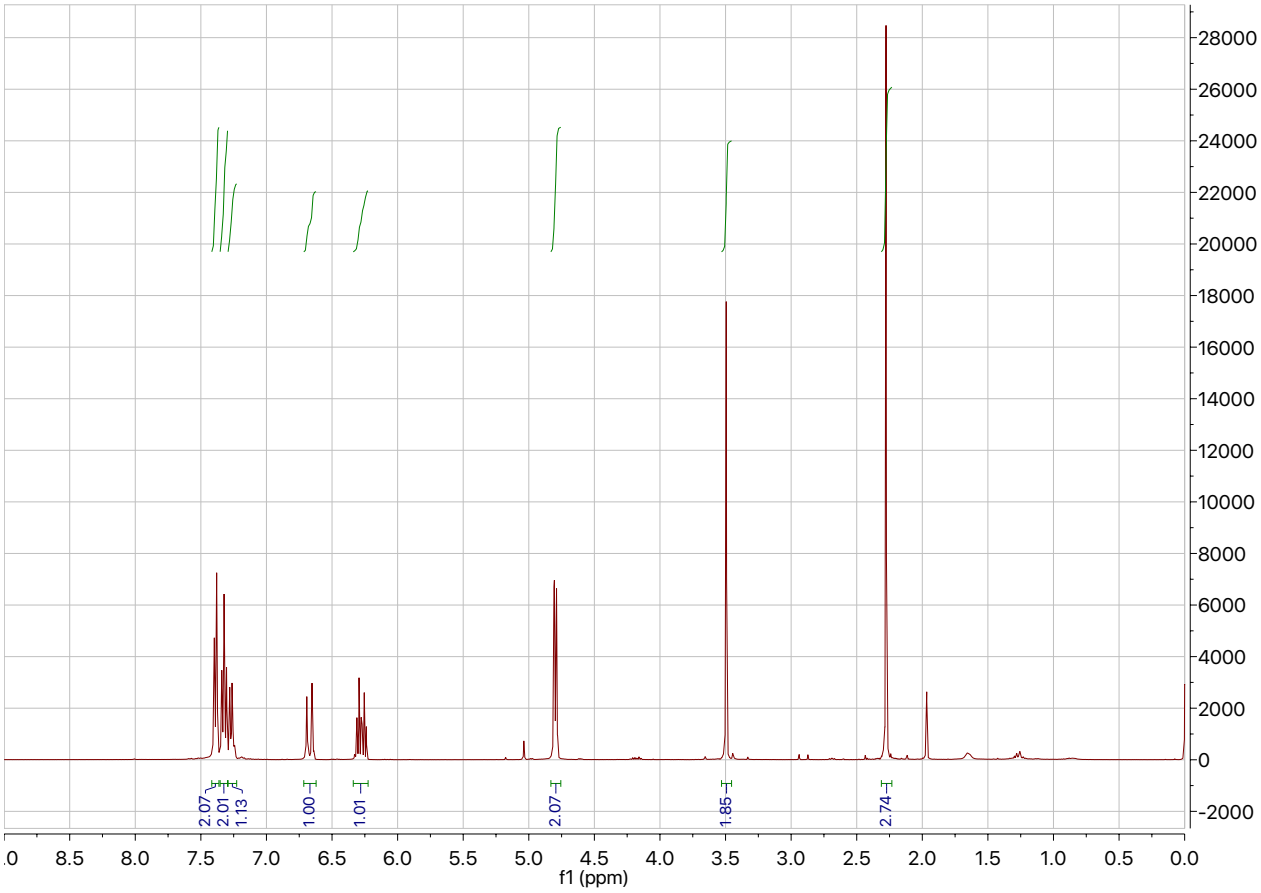


Figure S10. [1]H-NMR of compound **52**.

COMPOUND ***53***

The procedure was followed as reported by Beck *et al.*[325] A flame dried Schlenk tube was charged with $[Rh(COD)Cl]_2$ (4.9 mg, 10 µmol, 2.0 mol%) and DPEphos (16.2 mg, 30 µmol, 6.0 mol %). A premixed solution of degassed DCE and EtOH (5:1, 1.25 ml) was added. Then 1-phenyl-1propyne (0.75 mmol, 1.5 eq.), TFA (8.0 µl, 12 mg, 0.1 mmol, 20 mol %), and the 1,3-dicarbonyl compound (0.5 mmol. 1.0 eq.) were added. The tube was sealed and heated to 80 °C for 16 h. after Flash chromatography purification, the addition product was dissolved in ethanol (5 ml), powdered KOH (110 mg, 2.0 mmol, 4.0 eq.) was added and heated to 80 °C for 45 min. After cooling to room temperature, the reaction mixture was diluted with H_2O (20 ml), aq. sat. NH_4Cl-solution (10 ml) was added and the mixture was extracted with Et2O (3 × 25 ml). The combined organic layers were washed with brine (25 ml) and dried with anhydrous MgSO4. The solvent was removed under reduced pressure and the corresponding product was purified by Flash chromatography.

^{1}H NMR (500 MHz, Chloroform-d) δ 7.32 – 7.28 (m, 2H), 7.23 – 7.18 (m, 3H), 5.97 (ddd, 1H), 5.09 – 4.99 (m, 2H), 3.96 – 3.87 (m, 1H), 2.93 – 2.79 (m, 2H), 2.09 (s, 3H).

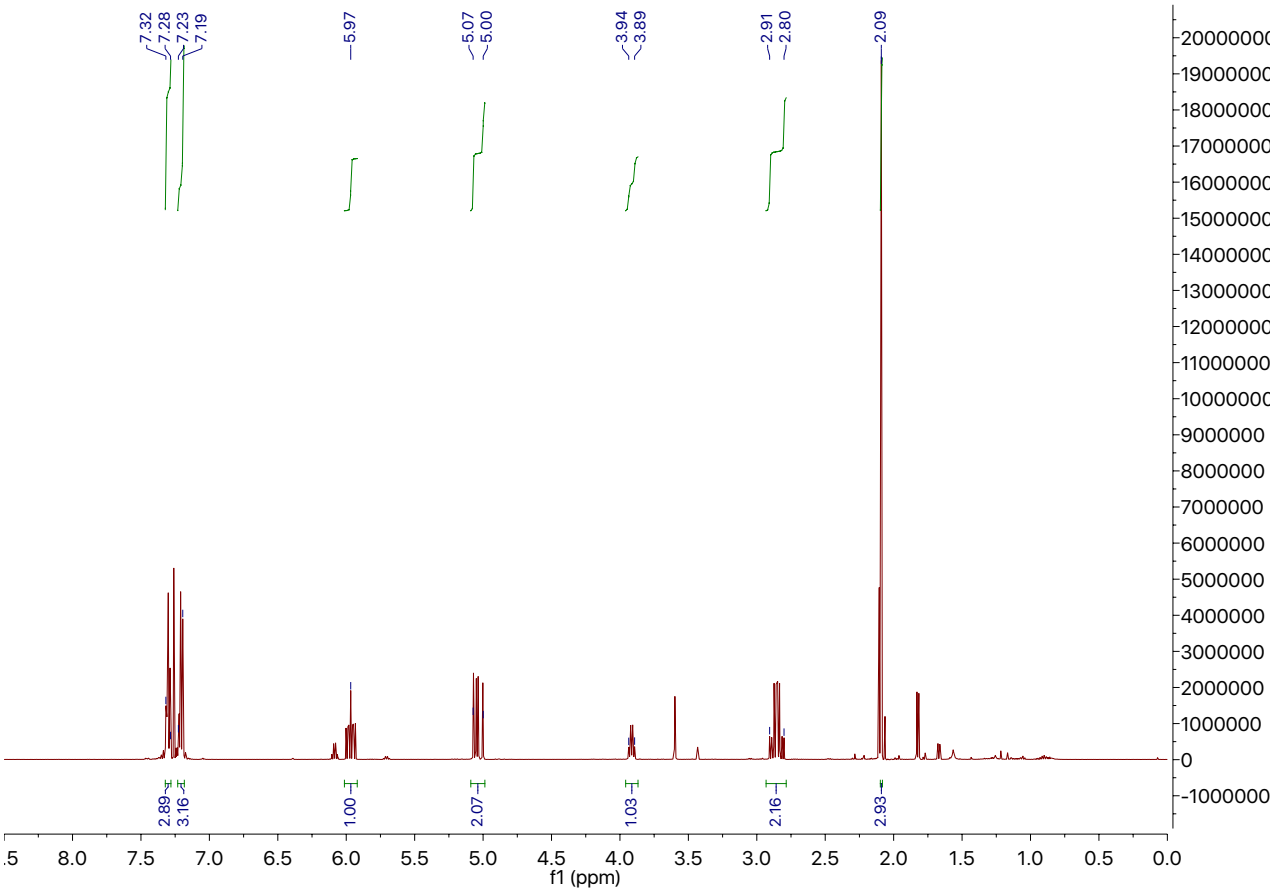


Figure S11. ^{1}H-NMR of compound **53**.

A.2.3 BUFFERS AND STOCK SOLUTIONS

Ligand stock solution	8 mM in DMF of the desired ligand.
Ru-precursor stock solution	$[CpRu(MeCN)_3]PF_6$, 8 mM in DMF.
assembled cofactor **46**	4 mM in DMF (the ligand stock solution and the cofactor stock solution were mixed in a 1:1 ratio)
Sav isoform	500 µM free biotin-binding sites in PBS-buffer
Substrate stock solution	40 mM substrate **52** in DMF
Product stock solution	20 mM product **53** in DMF
GSH stock solution	100 mM GSH in PBS
PBS buffer	50 mM NaH_2PO_4/Na_2HPO_4 (set to pH 7.4), 0.9 % NaCl

A.2.3 STANDARD LACOUR CONDITIONS

General procedure for the set-up of a Carroll rearrangement following Constant *et al.*

Briefly, in a 1 mL glass vial under N_2, the ruthenium precursor stock solution (in THF) and the ligand stock solution (in THF) were mixed in a 1:1 ratio to afford the CpRu-XX (10 mol% final concentration). The solution was incubated for 5 min at room temperature before the addition of the allyl-*β*-ketoesters **52** (1 mM final concentration). The reaction was incubated in a thermoshaker under inert atmosphere at 60 °C overnight.

A.2.5 General Procedure for Ligand Screening

In order to prepare the complexes, separate stock solutions of each ligand and the precursor $[CpRu(MeCN)_3]^+PF_6^-$ were prepared in dimethylformamide (DMF), each at 8 mM concentration. Ligands depicted in Scheme S1 were used for the screening. The ligands and the precursor stock solutions were subsequently mixed in a 1:1 ratio to yield 4 mM cofactor stock solutions.

200 µL test reactions were set up in a Teflon coated 96-well plate. Cofactor (20-200 µM final concentration) was dissolved in PBS (pH 7.5). Substrate (2 mM final concentration) was added. If GSH activity was screened for, GSH (10 mM final concentration) was added. The plate was incubated at 37 °C and 300 rpm.

Following, the product was extracted by the addition of 200 µL cyclohexane to each reaction well. The mixture was thoroughly vortexed and spun down in the centrifuge (4000 g). The organic phase was collected and analyzed by reverse-phase UPC^2 based on supercritical CO_2. The product concentration was determined by the UV absorption measurement at a wavelength of 210 nm. The UPC^2 was run with supercritical CO_2 as apolar phase and a heptane/isopropanol mixture (ratio 4:1) was applied as polar phase. The UPC^2 protocol is summarized in Table S3.

A.2.6 General Procedure for Sav-mutant Screening

The stock solutions were prepared similarly as described above. Complex **46** and complex **47** were used as cofactors.

200 µL test reactions were set up in a Teflon coated 96-well plate. Sav (400 µM final concentration) and cofactor (200 µM final concentration) were dissolved in PBS (pH 7.5) and incubated for 10-20 min at room temperature. Substrate (2 mM final concentration) was added and the plate was incubated at 37 °C and 300 rpm. Extraction of the product and analysis were carried out as described above.

A.2.7 UPC^2 Measurement

The UPC^2 was run with supercritical CO_2 as apolar phase and a heptane/isopropanol mixture (ratio 4:1) was applied as polar phase. The UPC^2 protocol is summarized in Table S3. The product (4.07 and 4.38 min) and substrate (2.75 min) were detected by UV absorption measurement at a wavelength of 210 nm and a temperature of 20 °C.

Table S3. UPC^2 protocol for the analysis of the Carroll rearrangement.

step	time [min]	flow [mL/min]	CO_2 [%]	heptane/iPrOH [%]
1	0.00	2.5	95	5.0
2	4.00	2.5	95	5.0
3	4.50	2.5	70	30
4	6.00	2.5	70	30
5	6.50	2.5	95	5.0
6	7.50	2.5	95	5.0

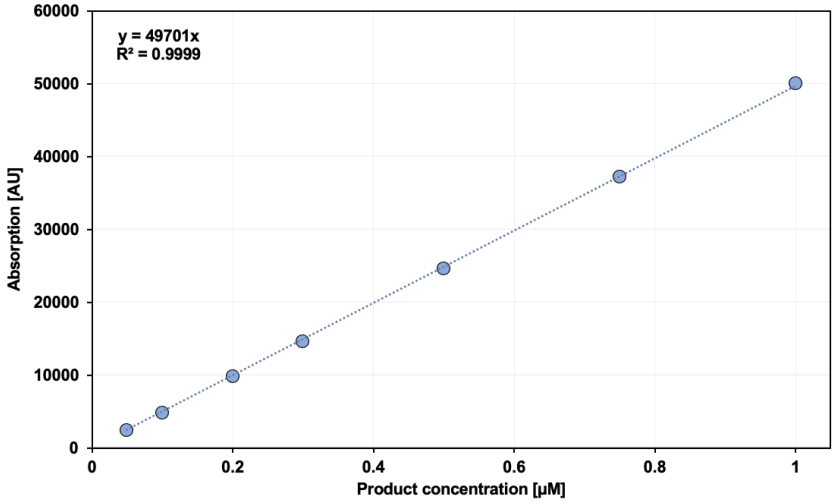


Figure S12. Carroll rearrangement product **53** calibration in the range 0.5-1 mM.

A.3 REFERENCES

[224] S. Constant, S. Tortoioli, J. Müller, J. Lacour, *Angew. Chem. Int. Ed* **2007**, *46*, 2082–2085.

[244] T. Heinisch, F. Schwizer, B. Garabedian, E. Csibra, M. Jeschek, J. Vallapurackal, V. B. Pinheiro, P. Marlì, S. Panke, T. R. Ward, *Chem. Sci.* **2018**, *9*, 5383–5388.

[211] T. Völker, F. Dempwolff, P. L. Graumann, E. Meggers, *Angew. Chem. Int. Ed. Engl.* **2014**, *53*, 10536–10540.

[325] T. M. Beck, B. Breit, *European J. Org. Chem.* **2016**, *2016*, 5839–5844.

APPENDIX B

B.1 SUPPLEMENTARY FIGURES AND TABLES

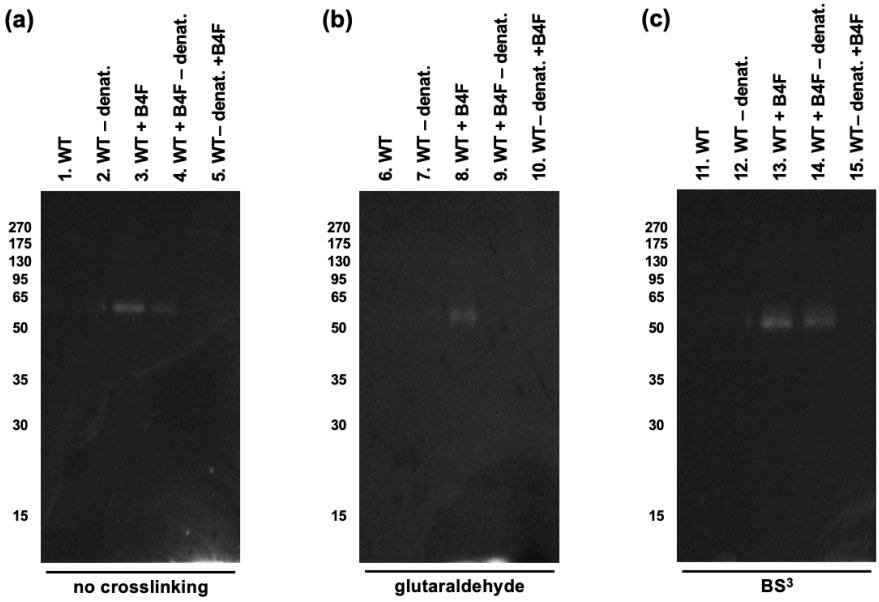


Figure S13. UV detection of SDS-PAGE (14% polyacrylamide gel) comparing different crosslinkers. a) Not crosslinked pp-Sav. **b)** Crosslinked with glutaraldehyde. **c)** Crosslinked with BS^3. 1) pp-Sav, 2) denatured after crosslinking, 3) pp-Sav incubated with B4F before crosslinking, 4) pp-Sav incubated with B4F, denatured and then crosslinked, 5) pp-Sav denatured first, then incubated with B4F and finally crosslinked. In the not crosslinked experiment, all samples were prepared the same way, just without the addition of a crosslinker.

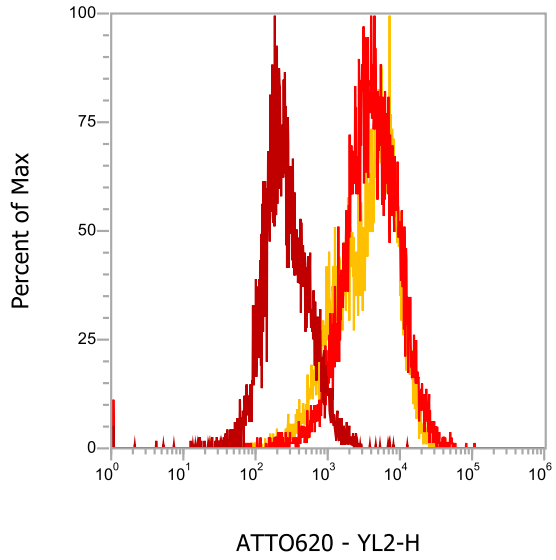


Figure S14. Expression level determined by staining the cells with a biotinylated ATTO-565 dye and analyzed by flow cytometry. dark red = empty vector, red = Sav^{SD}, yellow = Sav^{SUMO}.

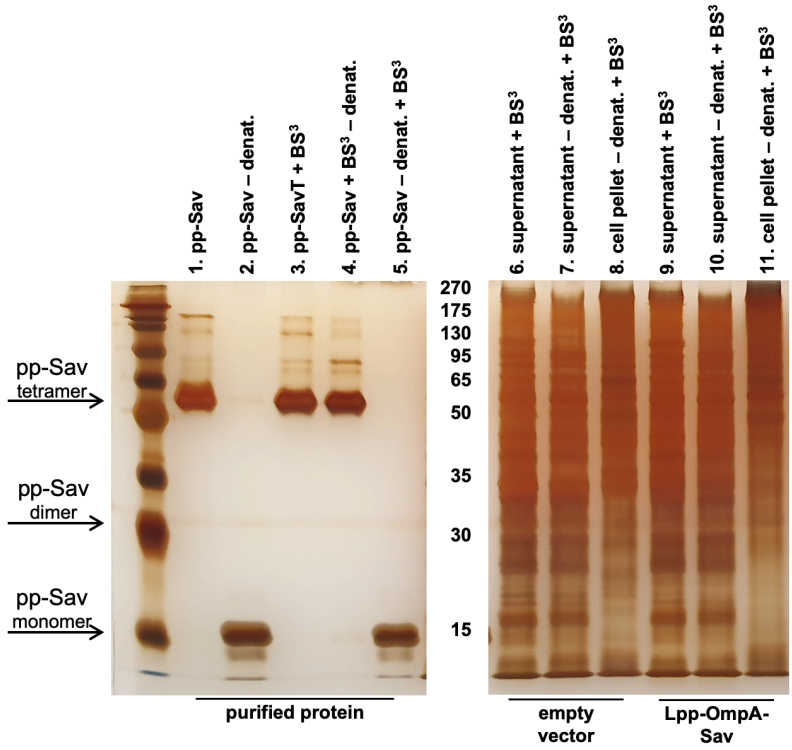


Figure S15. SDS-PAGE analysis of the quaternary structure of Sav^{SD}. Crosslinking of Sav^{SD}, followed by SDS-PAGE (14% polyacrylamide gel) and western blot (anti-Sav rabbit polyclonal antibody used at a 1:200 dilution). Except for lane 5, all samples were crosslinked with BS^3 prior to any other treatment. All denatured samples were treated under the same conditions: SDS sample buffer, 95°C, 30 min. 1) pp-Sav, 2) denatured pp-Sav, 3) crosslinked pp-Sav, 4) crosslinked and then denatured pp-Sav, 5) denatured and then crosslinked pp-Sav, 6) empty vector, supernatant, 7) empty vector, denatured supernatant, 8) empty vector, denatured cell pellet 9) Sav^{SD}, supernatant, 10) Sav^{SD}, denatured supernatant, 11) Sav^{SD}, denatured cell pellet.

B.2 EXPERIMENTAL PART

B.2.1 MATERIALS AND METHODS

All commercially available chemicals were purchased from Sigma-Aldrich, Acros Organics, Fluka, Fluorochem or Merck and used without further purification. Deuterated solvents were purchased from Cambridge Isotope Laboratories Inc. The water was purified with a Milli-Q-system (Millipore). Antibiotics were purchased from Applichem GmbH, DNase I was from Roche Diagnostics AG, IPTG was from Apollo Scientific, biotin-4-fluorescein was from ANAWA Trading SA and Agarose/SDS-PAGE markers were from New England BioLab Inc. Restriction enzymes, DNA polymerases and ligases were purchased from New England BioLab Inc. Primers were ordered from Microsynth AG. The double site saturation library was purchased from EUROFINS. Thin layer chromatography (TLC) was performed on Merck TLC Silica gel 60 F254 plates, using a UV-detector (254 nm or 360 nm). Basic $KMnO_4$ solution was used for the staining. Column chromatography was performed using silica gel (Merck Silica gel 60 (0.040-0.063 mm)). NMR spectra were measured on a 400 MHz and 500 MHz Bruker Advance spectrometer at room temperature and evaluated with MestReNova. Chemical shifts (δ) are quoted in parts per million (ppm) and referenced to the residual solvent peaks. Scalar coupling (J) is reported in Hertz (Hz). High resolution mass spectra (HRMS) were measured on a Bruker maXis 4G. Mass-spectral analysis of the expressed streptavidin mutants was performed on a Bruker Daltonics, ESI/micrOTOF MS. Fluorescence/Absorbance spectroscopy was performed on a TECAN infinite M1000 Pro. PCR reactions were performed with an Eppendorf Mastercycler Gradient. DNA sequencing (Sanger cycle sequencing/capillary electrophoresis) was performed by Microsynth AG. *E. coli* BL21(DE3) were purchased from Stratagene (Agilent), *E. coli* NEB5α cells were purchased from New England Biolabs and *E. coli* TOP10(DE3) cells were a gift from Dr. Markus Jeschek (DBSSE ETH Zürich, Switzerland). Chemically competent cells were prepared according to the Hanahan method using RbCl. Affinity column chromatography (purification of the expressed streptavidin mutants) was performed on an Äktaprime Plus chromatography system, using 2-iminobiotin sepharose column. Flow cytometry analysis was performed on an Attune NxT acoustic focusing cytometer (life technologies). High-throughput screening was performed with a Voaflo96 pipetting station from Integra.

B.2.3 Expression and purification of Streptavidin isoforms

Streptavidin isoforms were expressed in *E. coli* BL21(DE3) cells containing pET11b_SAV plasmids and purified as described elsewhere.[325] The number of free biotin-binding sites was determined by titration with biotin-4-fluorescein.[326]

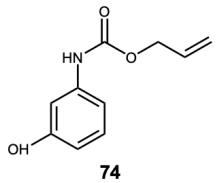
B.2.4 Synthesis

*Biotinylated ruthenium cofactor – cofactor **46***

The biotinylated ruthenium cofactor [CpRu(QA-Biot)(OH_2)] **46** was prepared *in situ* by mixing solutions of [CpRu($MeCN)_3$]PF_6 and biotinylated ligand in a 1:1 ratio in DMF and subsequent incubation for 10 min at room temperature. The biotinylated ligand was synthesized as described by Heinisch *et al*.[244]

*Caged Hydroxyaniline Substrate **74***

At 0°C, allyl chloroformate (3.90 ml, 4.42 g, 36.7 mmol, 2.0 eq.) was added dropwise to a stirred suspension of 3-aminophenol (**75**, 2.00 g, 18.3 mmol, 1.0 eq.) in a mixture of dry diethyl ether (100 ml) and dry THF (20 ml). A white precipitate (the amine hydrochloride) was formed immediately. The reaction mixture was stirred an additional 2 h at room temperature. The hydrochloride was removed by filtration. The filtrate was evaporated under reduced pressure. The obtained brown oil was solubilized in diethyl ether (20 ml) and the resulting organic phase was washed with solutions of HCl (1 M, 20 ml), saturated $NaHCO_3$ (20 ml) and water (20 ml). The organic phase was dried over Na_2SO_4 and the solvent was evaporated to yield a brown oil. The residue was purified by flash column chromatography (SiO_2, ethyl acetate/cyclohexane 1:3) to yield the product as a white crystalline solid (**74**, 792 mg, 4.1 mmol, 22% yield). NMR and HRMS spectra are illustrated in Figure S16.



74

Scheme S3. Compound 3

^{1}H NMR (500 MHz, Chloroform-*d* δ/ppm): 7.30 (s (broad), 1H), 7.13 (t, J = 8.1 Hz, 1H), 6.81 (s (broad), 1H), 6.68 (ddd, J = 8.0, 2.1, 0.9 Hz, 1H), 6.58 (ddd, J = 8.1, 2.4, 0.9 Hz, 1H), 6.52 (s (broad), 1H), 5.95 (ddt, J = 17.2, 10.4, 5.7 Hz, 1H), 5.36 (dq, J = 17.2, 1.5 Hz, 1H), 5.26 (dq, J = 10.4, 1.3 Hz, 1H), 4.67 (dt, J = 5.7, 1.4 Hz, 2H). Solvents: Chloroform (7.26), cyclohexane (1.43).

^{13}C NMR (126 MHz, Chloroform-*d* δ/ppm): 156.77 (1C), 153.62 (1C), 138.69 (1C), 132.07 (1C), 129.98 (1C), 118.54 (1C), 110.86 (1C), 110.58 (1C), 106.03 (1C), 66.17 (1C). Solvents: Chloroform (77.28, 77.03, 76.77).

HRMS (ESI-MS, pos.) m/z: $[M + H]^+$ calculated for $C_{10}H_{12}NO_3$: 194.0812, found: 194.0811.

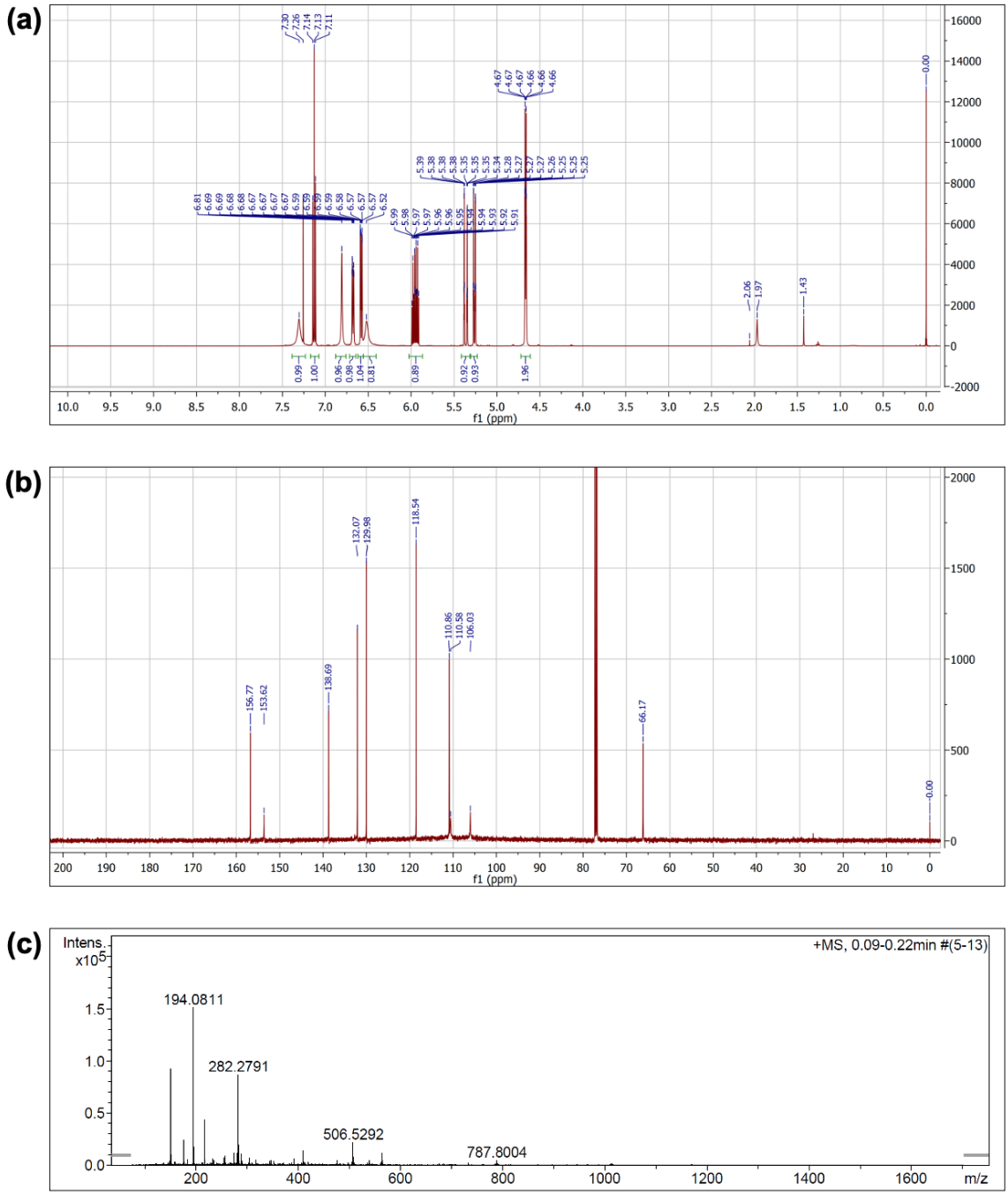


Figure S16. Characterization of compound **74**. **a)** ^{1}H-NMR, **b)** ^{13}C-NMR and **c)** HRMS spectrum.

B.2.5 CATALYSIS EXPERIMENTS

STOCK SOLUTIONS

Biotinylated ligand	5 mM in DMF
$[CpRu(MeCN)_3]PF_6$	5 mM in DMF
Ruthenium cofactor **46**	0.2 mM in DMF (4 µL of ligand **4** stock was mixed with 4 µL $[CpRu(MeCN)_3]PF_6$ stock and 92 µl DMF and incubated for 5 min)
Sav isoform	0.4 mM free biotin-binding sites in PBS-buffer
Substrate	50 mM substrate **3** in DMF
Product	20 mM product **2** in DMF
Internal standard	12.5 mM phenol in DMF
Inhibitor	10.4 mM potassium isocyanoacetate in water
PBS buffer	50 mM NaH_2PO_4/Na_2HPO_4 (set to pH 7.4), 0.9 % NaCl
LB-phosphate buffer	50 mM NaH_2PO_4/Na_2HPO_4, pH 7.4 (10.1 mL Na_2HPO_4 (200 mM), 2.4 mL NaH_2PO_4 (200 mM) and 37.5 mL LB medium were mixed together)

INDUCTION CAPACITIES OF VARIOUS PHENOLS AND ANILINES

E. coli DH5α cells containing the DmpR/GFP reporter plasmid were cultivated in LB-medium at 30°C to a cell density of OD_{600} of 1. The cell culture was diluted to OD_{600} = 0.05-0.08 in fresh media and aliquoted (1 mL) in fresh tubes Each tube was supplemented with 500 µM putative inducer and incubated at 30°C, 200 rpm overnight (12-13 h). The GFP fluorescence intensities and OD600 values of appropriately diluted overnight samples were recorded in a Tecan M2000 microtiter plate reader the OD-normalized fluorescence intensity values obtained with each putative inducer are show in Figure 2a.

IN VITRO CATALYSIS

Buffer (245 µL), protein (1.25 µL, 2 µM final concentration) and ruthenium cofactor **46** (1.25 µL, 1 µM final concentration) were added in this order to a 2 mL vial. The solution was incubated at room temperature for 10 min (formation of the ADAse). Substrate **3** (2.5 µL, 500 µM final concentration) was added (1.5% final DMF content). The vials were then closed and incubated in a thermoshaker (30°C, 18 h, 300 rpm shaking). The inhibitor was then added to the reaction (240 µL, 5 mM final concentration, followed by the internal standard (10 µL, 250 µM final concentration). No catalytic activity for ArM S112M-K121A spiked with inhibitor was detected[326]) followed by the internal standard. The sample was then subjected to HPLC analysis (Figure S18). The peak areas of the product and the internal standard in the HPLC chromatogram were integrated and the total turnover number (TON) was calculated by comparison to a standard curve (Figure S17)

IN VIVO CATALYSIS

In a 96-well plate deep well plate, 2.5 µL Sav (2 mM; 10 µM final concentration) were mixed with 2.5 µL Ru-cofactor (1 mM; 5 µM final concentration). The mixture was incubated for 5 min. Next, 25 µL substrate **3** (10 mM; 500 µM final concentration), 20 µL LB-phosphate and 450 µL reporter cells were added and the plate covered with a porous seal to start the reaction. The plate was incubated at 30 °C, 300 rpm. After 9 h, aliquots of the reaction were transferred to a TECAN plate reader to determine the GFP fluorescence (excitation: 488 nm, emission: 509 nm) and the OD_{600}.

B.2.6 EVALUATION BY HPLC

The HPLC measurement was run with mQ-H_2O + 0.3% MeCN + 0.1% TFA (solvent A) as polar phase and MeCN+0.1% TFA as apolar phase The HPLC protocol is summarised in Table S4. The product (1.82 min), substrate (11.77 min) and the internal standard (9.49 min) were detected by UV absorption measurement at a wavelength of 280 nm and an oven temperature of 40 °C. The column used was an Agilent Eclipse XDB-C18 5 µm 4.6x150mm.

Table S4. HPLC protocol for the analysis of the Carroll rearrangement.

step	time [min]	flow [mL/min]	solvent A [%]	solvent B [%]
1	0.00	1.0	95	10
2	5.00	1.0	95	10
3	7.00	1.0	70	30
4	15.0	1.0	70	30
5	17.0	1.0	95	10
6	22.0	1.0	95	20

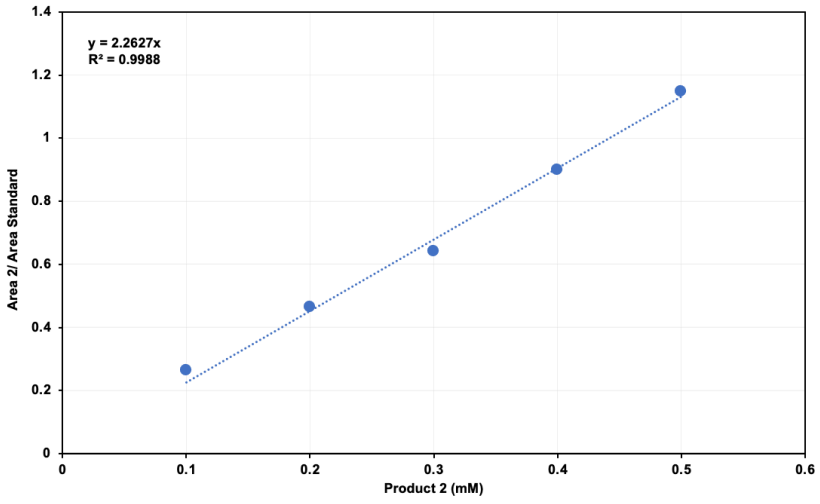


Figure S17. Standard curve of product **75** as a function of the peak areas of the product and the internal standard in the HPLC chromatogram.

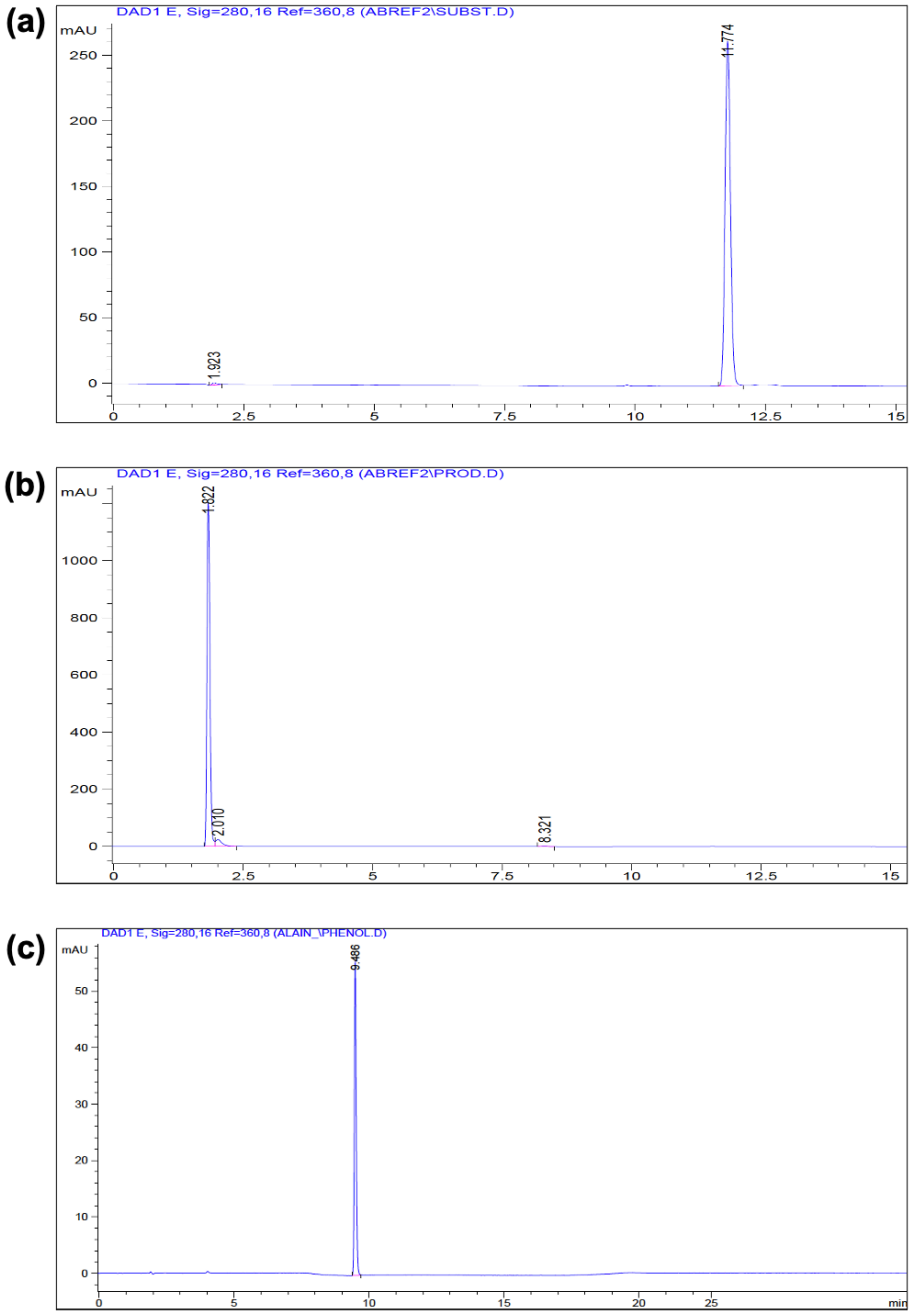


Figure S18. HPLC chromatogram (absorption at 280 nm) of **a)** Substrate **74**, **b)** Product **75** and **c)** Internal standard.

B.2.7 Design and Cloning of Streptavidin and Reporter Plasmids

Table S5. Plasmids used in this Study

Abbreviation	backbone	Gene	Inducer	Resistance marker
P1	pSEVA131	dmpR, sfGFP	3-OH-anillin	Ampicillin
P2	pBAD33	Lpp-OmpA-Sav	L-arabinose	Chloramphenicol

*Cloning of reporter plasmid reporter **P1***

A double-stranded synthetic DNA fragment encoding the *dmpR* gene (codon-optimized) from *Pseudomonas sp.* and promoter sequences P_r and P_o from *Pseudomonas sp.* was purchased from IDT DNA. The fragment was introduced in a modified pSEVA131 backbone (p15a ori, ampicillin antibiotic cassette, promoterless *sfgfp*) by Gibson assembly following manufacturer's protocol (NEB). The expression of the fluorescent protein is controlled by the P_o inducible promoter. Vector amplification was achieved by polymerase chain reaction (PCR) using the Phusion high-fidelity Polymerase with GC buffer, 5 mM dNTPs and 19 µM of each primer (5`-GGATC-CTAATTAATTAAAGGCATCAAATAAAACG) and (5`-CTTGGACTCCTGTTGATAGATC). The inserted sequence was sequence verified by Sanger sequencing at Microsynth AG, after plasmid isolation using the QIAquick Miniprep plasmid purification kit. Plasmid **P1** was transformed into chemically competent NEB5α (NEB) cells following the manufacturer´s protocol.

*Preparation of GFP reporter cells (NEB5α-**P1**)*

NEB5α_pDmpR cells were freshly picked from an LB-agar$_{amp}$ plate or from a glycerol stock to inoculate LB_{amp} medium for preparation of an overnight culture. In the next morning, 20 mL LB_{amp} medium were inoculated with 600 µL overnight culture in a 250 mL baffled shake flask and incubated at 37 °C, 250 rpm, until ~OD_{600} of ~0.6 was obtained (~2.5 h). Prior to the catalysis experiments the GFP reporter cells were stored on ice.

*Cloning of Sav display plasmid **P2***

For streptavidin expression and anchoring to the outer membrane, a fusion construct consisting of the *E. coli* lipoprotein (Lpp), a truncated version of the *E. coli* outer membrane protein A (OmpA) and full-length streptavidin (Sav) was cloned into an *L*-arabinose inducible pBAD33 vector backbone as described by Heinisch and Schwizer *et al.*[179]

*PREPARATION OF CATALYTICALLY COMPETENT GFP REPORTER CELLS (TOP10(DE3)-**P1-P2**)*

Plasmids **P1** (100 ng) and **P2** (100 ng) were co-transformed into 50 µL chemically competent Top10(DE3) by heat shock treatment (42 °C) and cells were incubated on LB-agar$_{cam,amp}$ plates (35 µg/mL chloramphenicol, cam; 50 µg/mL ampicillin, amp) at 37 °C overnight.

DOUBLE SITE SATURATION MUTAGENESIS LIBRARY GENERATION

A double site saturation mutagenesis library of Sav including positions S112 and K121 was selected for the optimization of the ADAse displayed on the *E. coli* surface. A two-step process was anticipated as a suitable library generation strategy:

(i) generation of a Sav gblock of a 90 bp length compromising DNK codons in positions S112 and K121. The gblock was purchased from the company EUROFINS.

(ii) Gibson assembly of the **P2** plasmid backbone and the Sav gblock library.

To reduce the screening effort, a focused Sav gblock library was designed containing degenerate codons DNK in both positions S112 and K121. The degenerate codon DNK codes for all 20 amino acids except Gln, Pro and His. Gibson assembly was carried out with the P1 backbone (obtained by PCR amplification with primers (5`- CAACCACCACGTCTTCGTGCGTAATTATGGGTCACC) and (5`- GTTGGTCATGATACCTTTACCAAAGTTAAACCGAGCGC) and the Sav gblock library to obtain the library for the in vivo screening.

DNA SEQUENCE OF SAV

turquoise = gblock library purchased from Eurofins, red = positions S112 and K121.

catgaccggtggccagcagatgggtcgtgatcaggcaggtattaccggcacctggtataatcagctgggtagcacctttattgttacc
gcaggcgcagatggtgcactgaccggtacgtatgaaagcgcagttggtaatgcagaaagccgttatgttctgacaggtcgttatgatagcgca
ccggcaaccgatggtagcggcaccgcactgggttggaccgttgcatggaaaaataactatcgtaatgcacatagcgcaaccacctggtcag
gtcagtatgttggtggtgcagaagcacgcattaatacccagtggctgctgaccagcggcaccaccgaagcaaatgcctggaaaagcaccct
ggttggtcatgataccttaccaaagttaaaccgagcgcagcaagcattgatgcagcaaaaaaagccggtgtgaataatggtaatccgctgga
tgcagttcagcag

LIBRARY PREPARATION BY GIBSON CLONING

After the PCR amplification using the primers (5`- CAACCACCACGTCTTCGTGCGTAA-TTATGGGTCACC) and (5`- GTTGGTCATGATACCTTTACCAAAGTTAAACCGAGCGC), 0.5 µL of DpnI was added to the sample in order to digest the template P2. The mixture was incubated at 37°C for 2 h followed by incubation at 80°C for 20 min (heat inactivation of the restriction enzyme). The sample was then purified following the manufacturer´s protocol

(Macherey-Nagel Nucleo Spin PCR clean up). For a final amount of ~4000 clones 4x 20 µL Gibson reaction were set up in parallel using a 5-fold excess of the Sav gblock library (insert, 0.15 pmol final amount per 20 µL reaction) *vs.* the PCR amplified **P2** backbone (0.03 pmol final amount per 20 µL reaction). The Gibson reaction was carried out following the manufacturer´s protocol (New England Biolabs; incubation at 50°C for 60 min).

*TRANSFORMATION OF SAV LIBRARY PLASMID P2 AND GFP REPORTER PLASMID **P1***

The transformation procedure was carried out in two steps. First, the Gibson-assembled Sav library plasmid **P2** was transformed into *E. coli* DH5α cells and afterwards isolated. In parallel, chemically competent *E. coli* Top10(DE3) cells which already contain the GFP reporter plasmid **P1** were prepared. Finally, the isolated **P2** plasmid was transformed into the Top10(DE3) cells containing **P1**.

A volume of 2 x 5 µL Gibson reaction mixture was transformed with 2 x 50 µL chemically ultra-competent *E. coli* DH5α cells (New England Biolabs) using a heat shock method (42 °C). Afterwards, the cells were taken up in 2 x 1 mL LB-medium (containing 34 µg/ml chloramphenicol) and 2 x 50 µL thereof were transferred onto LB-agar$_{cam}$ plates and incubated at 37 °C overnight. The remaining 2 x 950 µL were incubated overnight (37 °C, 250 rpm). On the two LB-agar plates ~2500 colonies were counted, which corresponds to a total number of ~47000 individual clones present in the liquid culture. The plasmids were then isolated following the manufacturer´s protocol (Macherey-Nagel Nucleo Spin Plasmid kit). A minimal amount 1 ng of the isolated plasmid was transformed into 50 µL chemically competent *E. coli* TOP10(DE3) cells which already contain the **P1** plasmid using a heat shock method (42 °C). In this way it should be guaranteed that the final cells only contain one individual streptavidin clone per cell. Afterwards, the cells were taken up in 1 mL LB-medium (containing 34 µg/mL chloramphenicol and 100 µg/mL ampicillin) and 20 µL thereof were transferred onto LB-agar$_{Cam,\ Amp}$ plates and incubated at 37 °C overnight. The remaining 980 µL were incubated overnight (37 °C, 250 rpm). On the LB-agar plate ~70 colonies were counted, which corresponds to a total number of ~3400 individual clones present in the liquid culture.

SEQUENCING AND EVALUATION OF THE LIBRARY

2 x 96 colonies from the final library (section 0) were sequenced and the amino acid distributions at the positions S112 and K121 were evaluated (Figure S19). 58% of the sequenced colonies were desired single or double mutants, 12% were the initially used template (K121A), 17% contained a frame shift within the gene, 4% contained an unwanted stop codon or an insert and 9%

of the sequencing gave unclear results. We applied the following equation to calculate the number of transformants which need to be tested in order to obtain a certain coverage of the library (Equation S1).[327]

$$T = -V * \frac{1}{S} * \ln(1 - P_i)$$

Equation S1. Calculation of the number of colonies to be screened to reach a certain coverage of the library. T: Number of transformants to be tested, V: Number of variants within the library, S: Percentage of correct sequences within the library, Pi: Percentage of coverage.

Two degenerate DNK codons at the positions S112 and K121 lead to a total number of variants within the library of V = 24^2 = 576. With S = 58% of correct sequences in the library and a total number of T = 2762 analyzed transformants, we obtained a percentage of coverage of P_i = 93.8%.

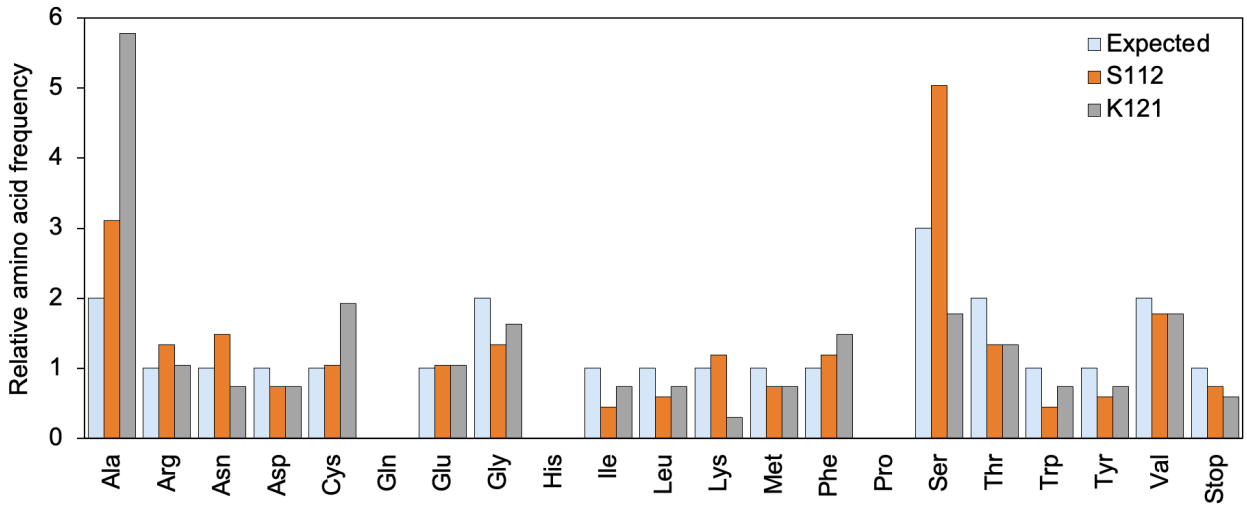


Figure S19. Amino acid distribution of the prepared streptavidin double site saturation library at positions S112 and K121. The relative amino acid frequencies calculated from 162 sequenced colonies are displayed. The expected amino acid frequencies were calculated based on the degenerate codon DNK.

Table S6. General PCR conditions.

Reagents	**1 PCR reaction (µL)**
Q5 buffer, 5x	5
dNTPs (10 mM)	0.5
Gblock_backbone_fw (10 µM)	0.5
Gblock_backbone_rv (10 µM)	0.5
Plasmid prep template **P2** (25 ng/µL)	1
Q5 Hot Start DNA polymerase (2 U/µL)	0.25
ddH_2O	15.25
DMSO (8% final)	2
Total volume	25

Table S7. General PCR thermocycling protocol.

Initial denaturation		95 °C	2 min
25 cycles	Denaturation	95 °C	30 s
	Annealing	70 °C	15 s
	Extension	72 °C	5 min
Elongation		72 °C	10 min
Hold		8 °C	

B.2.9 HIGH THROUGHPUT SCREENING OF SURFACE-DISPLAYED ADASES

The 96-well plate assay was carried out following a modified protocol described by Jeschek *et al*.[187]

STOCK SOLUTIONS

Biotinylated ligand	5 mM in DMF
$[CpRu(MeCN)_3]PF_6$	5 mM in DMF
Ruthenium cofactor **46**	2 mM in DMF (e.g. 60 µL of ligand **4** stock was mixed with 60 µL $[CpRu(MeCN)_3]PF_6$ stock and 30 µl DMF and incubated for 5 min at room temperature)
Cofactor **46** buffer	2 µM ruthenium cofactor **46** in PBS buffer
Substrate buffer	500 µM substrate **3** in modified Studier 5052 medium
Atto 565-Biotin buffer	2 µM Atto 565-Biotin in PBS buffer
PBS buffer	50 mM NaH_2PO_4/Na_2HPO_4 (set to pH 7.4), 0.9 % NaCl
LB medium	10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl, 100 µg/mL ampicillin, 34 µg/mL chloramphenicol
Modified Studier 5052 medium	25 mM Na_2HPO_4, 25 mM KH_2PO_4, 50 mM NH_4Cl, 5 mM Na_2SO_4, 2 mM $MgSO_4$, 0.5% glycerol, 15 g/L tryptone and 10 g/L yeast extract, 100 µg/mL ampicillin, 34 µg/mL chloramphenicol
LB-phosphate buffer	50 mM NaH_2PO_4/Na_2HPO_4, pH 7.4 (10.1 mL Na_2HPO_4 (200 mM), 2.4 mL NaH_2PO_4 (200 mM) and 37.5 mL LB medium were mixed together)

CELL CULTURES AND PROTEIN EXPRESSION:

All pipetting steps were performed with a 96-well pipetting station from Integra. In a 96-flat well plate, 120 µL of LB-medium were inoculated with fresh colonies of Sav cells (Top10(DE3): pBAD33_Lpp-OmpA-Sav + pDmpR_GFP). Each 96-flat well plate contained 90 colonies and 6 control samples (3x: empty cells; 3x: Sav S112M-K121A). After incubation overnight (37 °C, 200 rpm), 100 µL of LB-medium were added to the preculture plate. 300 µL of modified Studier 5052 medium were filled into a fresh 96-deep well plate and then inoculated with 10 µL preculture. The main cultures were incubated at 37 °C for 3.75 h at 380 rpm followed by addition of 30 µL of a 2 % L-arabinose solution (0.2 % final concentration) and incubation for 4 h at 25 °C and 380 rpm. The optical density OD_{600} of the culture after streptavidin expression was determined to be 7 – 8.

COFACTOR LOADING:

The plate was centrifuged (5 min, 10°C, 3.200 g), the supernatant was discarded and the pellets were resuspended in 600 µL ice-cold PBS buffer. In case of a subsequent determination of the streptavidin expression level, an aliquot of 50 µL was transferred into a new 96-deep well plate (only done for the final hits). The plate was centrifuged (5 min, 10°C, 3.200 g), the supernatant was discarded and the pellets were resuspended in 600 µL cofactor buffer. The cells were incubated on ice for 30 min. Finally, the plate was centrifuged (5 min, 10°C, 3.200 g), the supernatant was discarded and the cell pellets were washed 2 x in 600 µL ice-cold PBS buffer.

CATALYSIS:

The pellets were resuspended in 200 µL substrate buffer (or modified Studier 5052 medium containing a defined amount of product **45**), the 96-deep well plate covered with a porous cover slide and incubated at 30 °C, 380 rpm for 16 h. The optical density OD_{600} of the culture at the beginning of the catalysis was determined to be 11-12. After the catalysis, 10 µL of the reaction mixture were mixed with 190 µL PBS buffer in a transparent 96-flat well plate. The optical density OD_{600} of the cells as well as the GFP fluorescence (excitation = 475 nm, emission = 509 nm) were measured in a TECAN plate reader.

DETERMINATION OF THE RELATIVE SAV EXPRESSION LEVELS:

The plate containing the 50 µL aliquots (from section 0) was centrifuged (5 min, 10°C, 3.200 g), the supernatant was discarded and the pellets were resuspended in 50 µL Atto 565-Biotin buffer. The cells were incubated on ice for 30 min. The plate was centrifuged (5 min, 10°C, 3.200 g), the supernatant was discarded and the cell pellets were washed 2 x in 600 µL ice-cold PBS buffer. The pellets were resuspended in 200 µL PBS buffer and 150 µL of the cell suspension was transferred into a flat 96-well plate. The Atto 565 fluorescence of cells was measured in a plate reader (TECAN, excitation at 565 nm and emission at 590 nm, Figure S20).

EVALUATION OF THE SCREENING RESULTS

The variation of ADAse activity/GFP expression level of a 96-deep well plate containing 96 times the Sav WT ADAse was determined to be 16.1%. No distinct dependency of the GFP expression level on the position within the 96-well plate was observed.

The 10 best hits from the library screening (K121D, K121E, K121G, K121N, S112M-K121R, S112M-K121T, S112T-K121A, S112T-K121G, S112T-K121N, S112T-K121T), the according

single mutants (S112M, S112T, K121A, K121R, K121T) as well as the wild-type (Sav WT), the mutant S112M-K121A and the empty vector control were selected and rescreened.

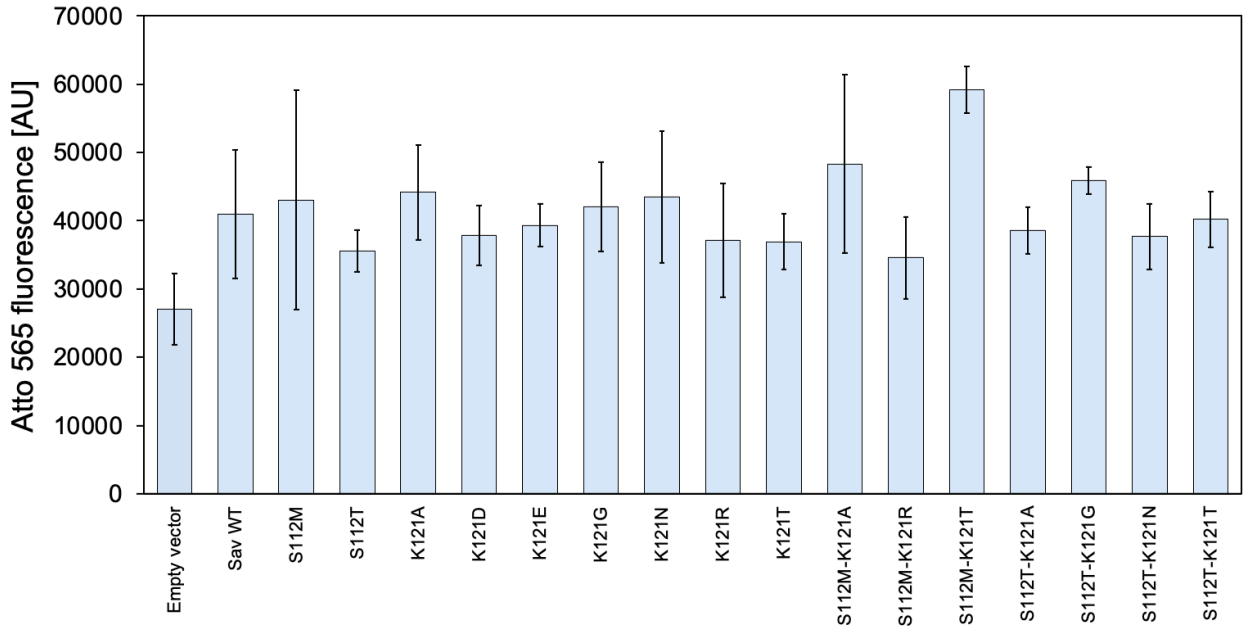


Figure S20. Determination of the relative Sav expression levels on the surface of *E. coli* cells. The relative Sav concentrations on the *E. coli* cell surface were determined via labelling with the biotinylated Atto 565-Biotin dye (section 0). Displayed are the mean values of the Atto 565-fluorescence (excitation = 565 nm, emission = 590 nm) ± 1 standard deviation of a measurement with eight repetitions. Displayed are the mean values of the GFP-fluorescence (excitation = 475 nm, emission = 509 nm) ± 1 standard deviation of a measurement with eight repetitions. For comparison, four cultures expressing Sav WT were spiked with increasing amounts of product.

B.2.10 Crosslinking Experiments

Preparation of the crosslinked samples

Precultures of *E. coli* TOP10(DE3) cells containing the plasmid for the Lpp-OmpA-Sav, and the empty vector, were prepared in LB-medium (34 µg/ml chloramphenicol) and incubated overnight at 37°C and 300 rpm in a 15 mL culturing tube. The modified Studier5052 medium (25 mL, 34 µg/ml chloramphenicol) was inoculated to an OD_{600} of 0.05 in a 250 mL baffled shake flask. The main culture was incubated (37°C, 300 rpm) until an OD_{600} of 0.4 – 0.8 was obtained (~2 h). Sav expression was induced by addition of 0.2% L-arabinose and Sav was expressed for 24 h at 25°C and 300 rpm.

Cells (OD_{600} 5/mL) from the cultivation were transferred into Eppendorf tubes, spun down and washed with ice-cold PBS (2x 500 µL). The pellets were then resuspended in 100 µL PBS and crosslinked with BS^3 (40 µL from a 25 mM stock solution, 8 mM final concentration) and incubated at room temperature for 30 min before the crosslinking was terminated with Tris-buffer (1 M, pH 7, 150 µL). The cell suspension was clarified and the supernatant discarded. Following, B-PER (200 µL) was added to the pellet and incubated for 1h at 30 °C. After incubation, the samples were

clarified by centrifugation (21'100 g, 15 min). The supernatant was separated and 10 ul were directly taken for analysis (sample 1). 10 µL were denatured in SDS sample buffer (6x Lämmli-buffer, 95 °C for 30-45 min) to give sample 2. The remaining pellet was resuspended in urea (200 µL, 8 M in 20 mM Tris-HCl buffer) and incubated for 30 min at 37°C. After incubation, the samples were clarified by centrifugation (21'100 g, 15 min) and the supernatant was denatured (6x Lämmli-buffer, 95 °C for 30-45 min) to give sample 3.

PURIFIED PROTEIN CONTROLS

All control samples are prepared from purified and lyophilized streptavidin WT (pp-WT). A stock solution of pp-WT (2 mg/mL) was prepared in PBS. For samples **2-6**, 10 µL of the stock solution was crosslinked with BS^3 (4 µL from a 25 mM stock solution, 10 mM final concentration) and incubated on ice (5 min). The reaction was terminated with Tris-buffer (1 M, pH 7, 10 µL). If B4F was added, B4F (2 µL) and Lämmli-buffer (20 µL, 6x conc.) otherwise only Lämmli-buffer (20 µL, 6x conc.) were added. If the probe was denatured, the sample was incubated at 95 °C for 30-45 min prior to the loading on the gel. Sample **5** was first denatured, then crosslinked and B4F added.

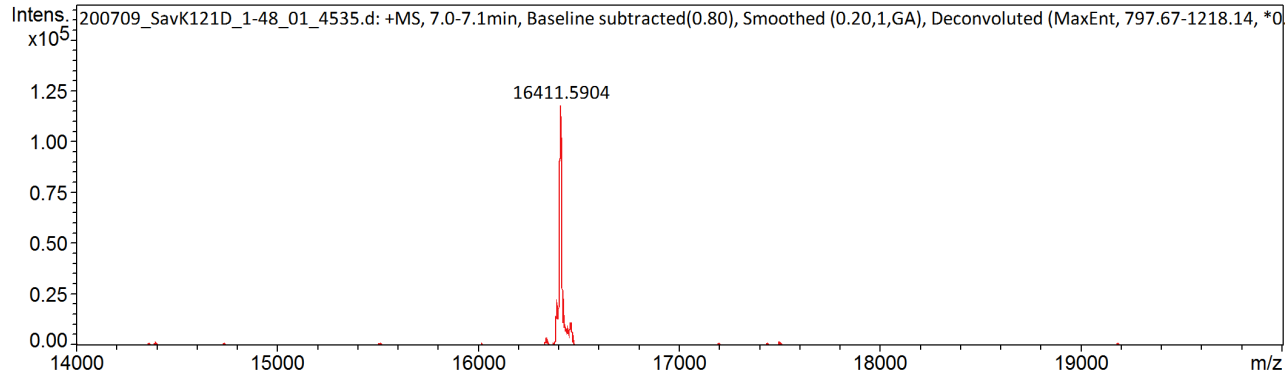
SDS-PAGE AND WESTERN BLOT

Two identical gels were loaded with 10 µL of each prepared sample and then run at 140 V for ~50 min. After completion, one gel was stained with a silver staining kit according to manufacturer's protocol (ThermoFisher Scientific).

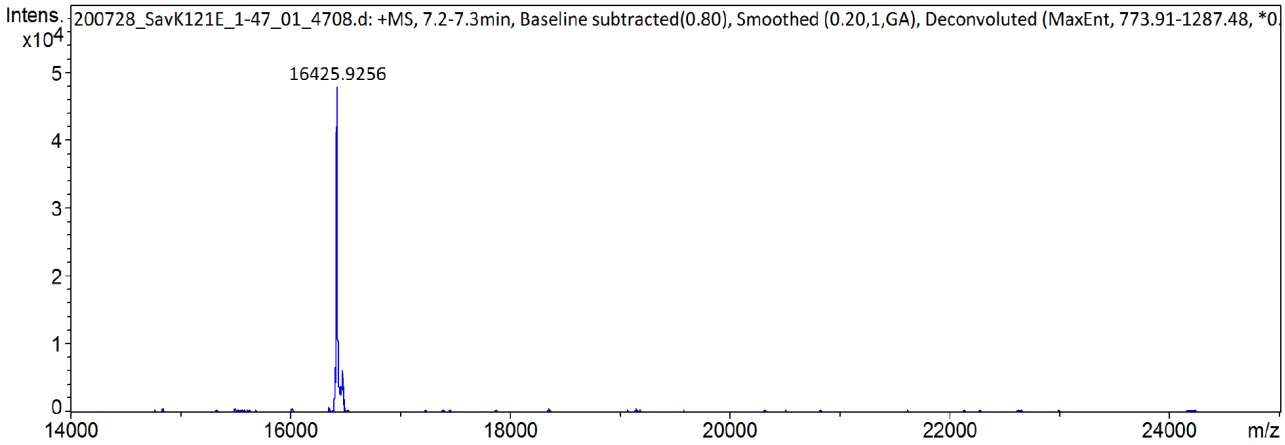
The second gel was transferred to a nitrocellulose membrane at 45 V overnight using a BioRad transfer system. After completion of the transfer, the membrane was quickly rinsed with water and blocked with TBST containing 4% BSA (25 mL) at RT for 1.5 h. After rinsing the membrane twice with TBST the membrane was incubated in fresh TBST (20 mL) containing the primary antibody (100 uL of a polyclonal serum, anti-Sav rabbit). After thorough washing with TBST (4 x 10 min) the membrane was incubated in fresh TBST (20 mL) containing the secondary antibody (10 uL anti-rabbit) fused with HRP. After washing the membrane with TBST (4 x 10 min) the detection was carried out according to manufacturer's protocol (ThermoFisher Scientific).

B.2.11 Deconvoluted Mass Spectra of the Purified Proteins

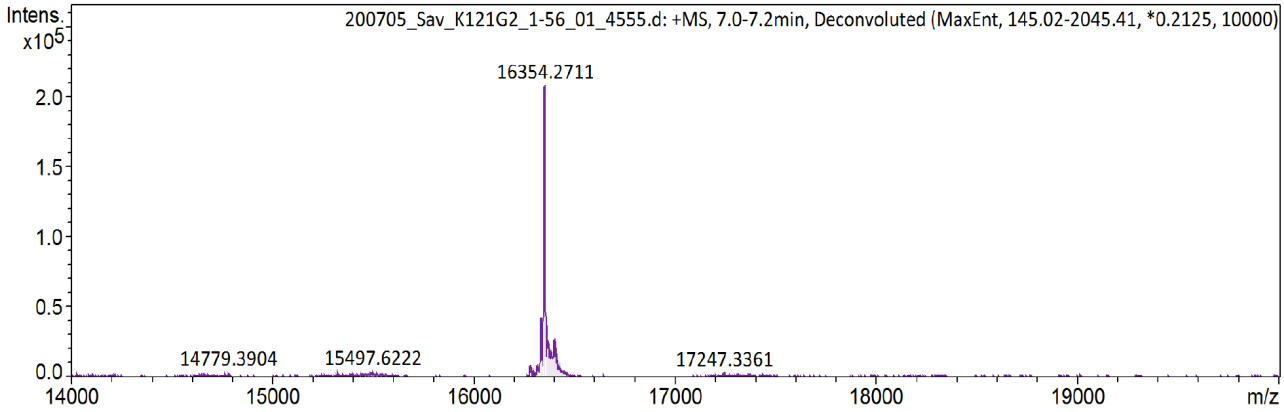
K121D (expected MW: 16412.82 g/mol)



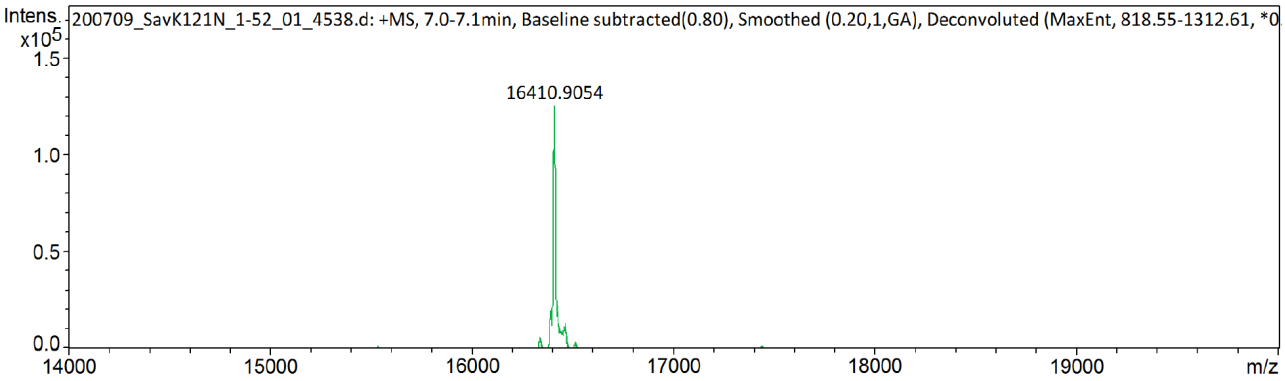
K121E (expected MW: 16426.84 g/mol)



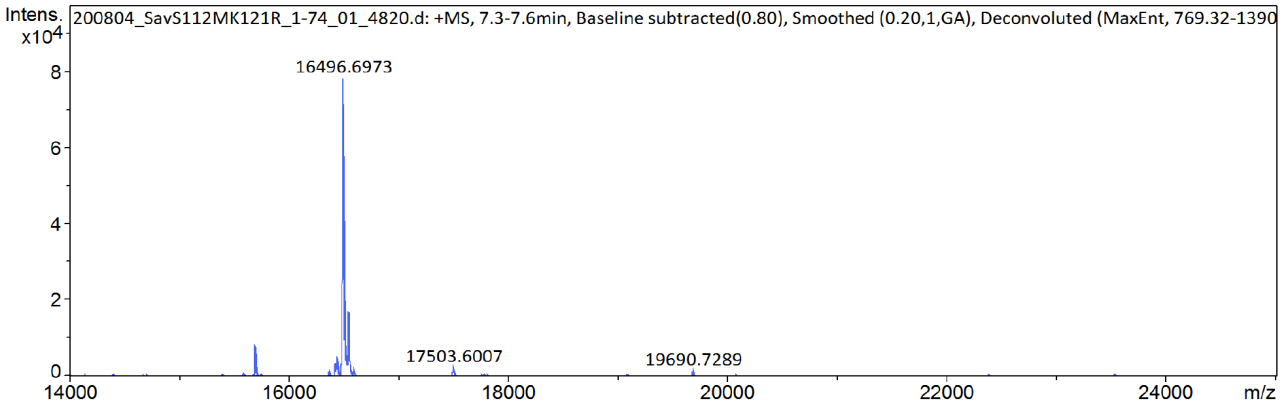
K121G (expected MW: 16354.78 g/mol)



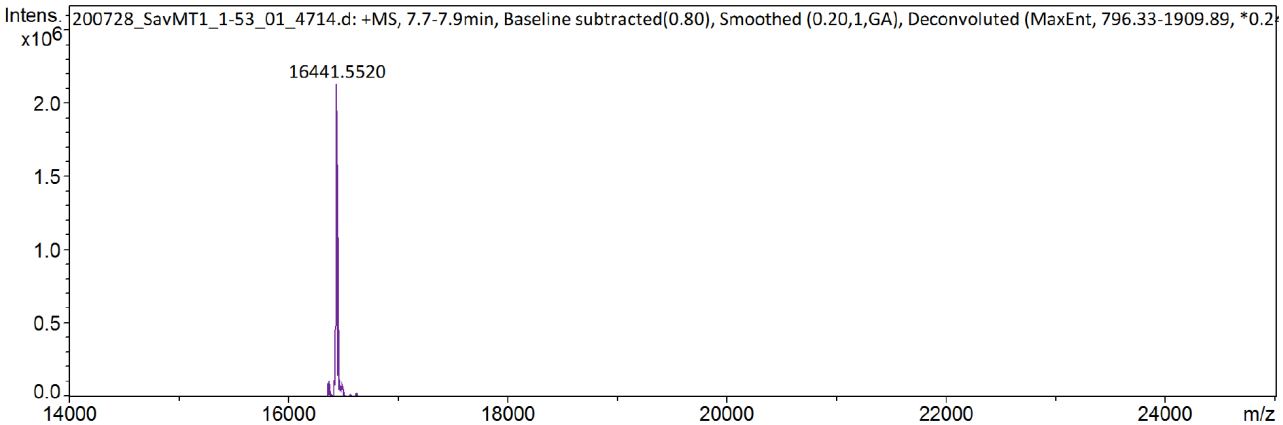
K121N (expected MW: 16411.83 g/mol)



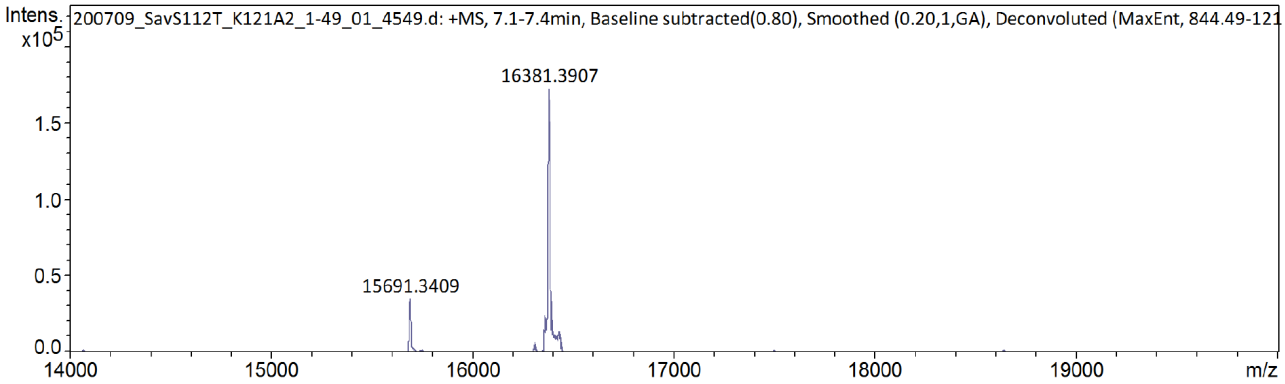
S112M-K121R (expected MW: 16498.03 g/mol)



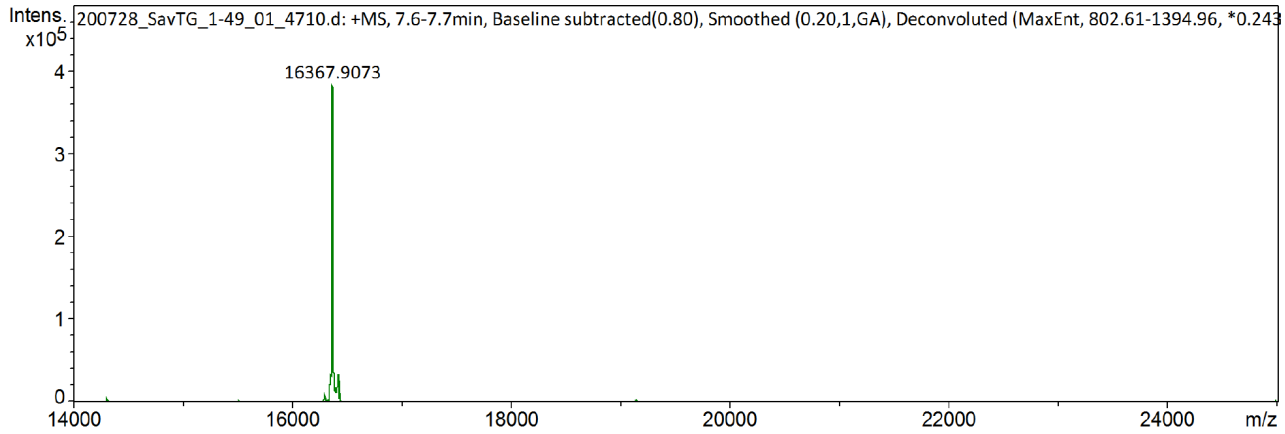
S112M-K121T (expected MW: 16442.95 g/mol)



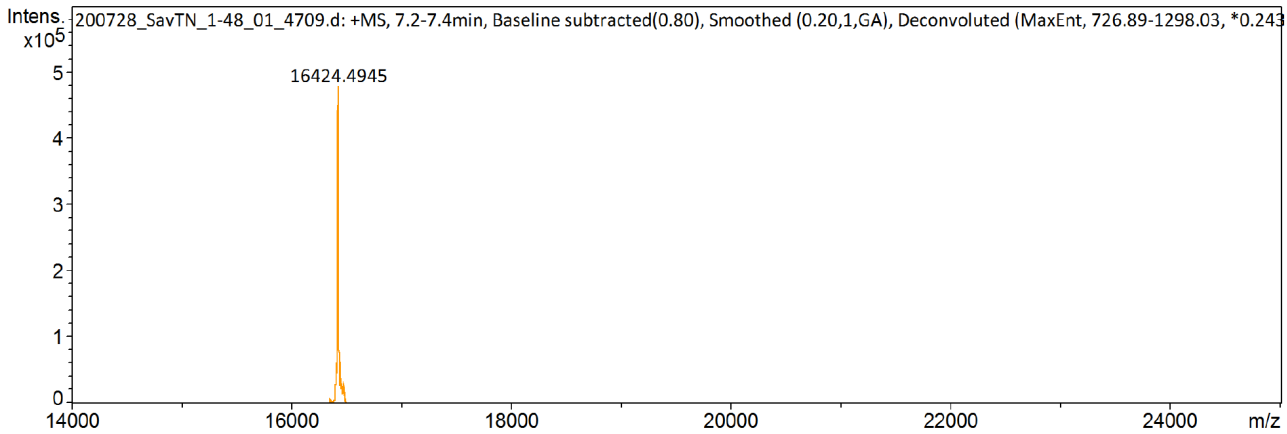
S112T-K121A (expected MW: 16382.83 g/mol)



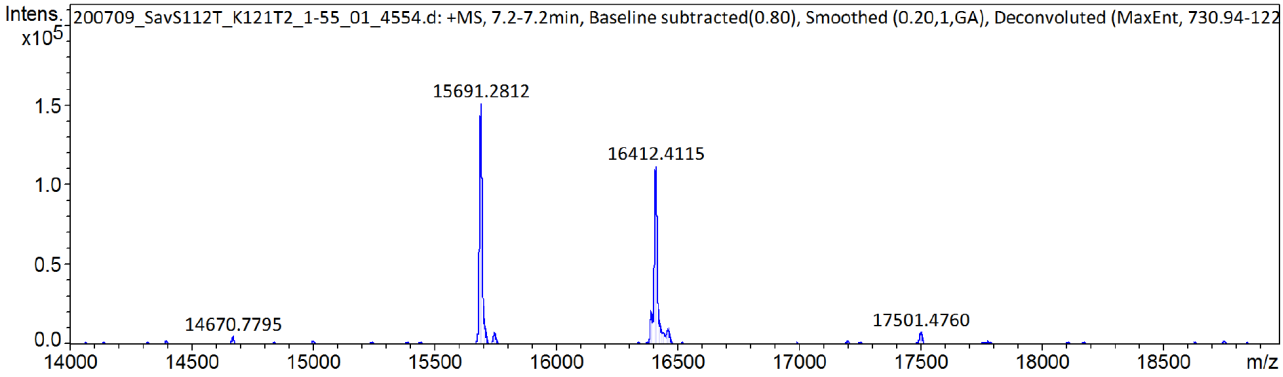
S112T-K121G (expected MW: 16368.81 g/mol)



S112T-K121N (expected MW: 16425.86 g/mol)



S112T-K121T (expected MW: 16412.86 g/mol)



B.3 REFERENCES

[325] V. Köhler, J. Mao, T. Heinisch, A. Pordea, A. Sardo, Y. M. Wilson, L. Knörr, M. Creus, J. C. Prost, T. Schirmer, T. R. Ward, *Angew. Chemie - Int. Ed.* **2011**, *50*, 10863–10866.

[326] G. Kada, H. Falk, H. J. Gruber, *Biochim. Biophys. Acta - Gen. Subj.* **1999**, *1427*, 33–43.

[244] T. Heinisch, F. Schwizer, B. Garabedian, E. Csibra, M. Jeschek, J. Vallapurackal, V. B. Pinheiro, P. Marlière, S. Panke, T. R. Ward, *Chem. Sci.* **2018**, *9*, 5383–5388.

[326] Y. Okamoto, R. Kojima, F. Schwizer, E. Bartolami, T. Heinisch, S. Matile, M. Fussenegger, T. R. Ward, *Nat. Commun.* **2018**, *9*, 1–7.

[327] M. T. Reetz, D. Kahakeaw, R. Lohmer, *ChemBioChem* **2008**, *9*, 1797–1804.

[187] M. Jeschek, R. Reuter, T. Heinisch, C. Trindler, J. Klehr, S. Panke, T. R. Ward, *Nature* **2016**, *537*, 661–665.

APPENDIX C

C.1 SUPPLEMENTARY FIGURES AND TABLES

C1.1 AGAR PLATE ASSAY

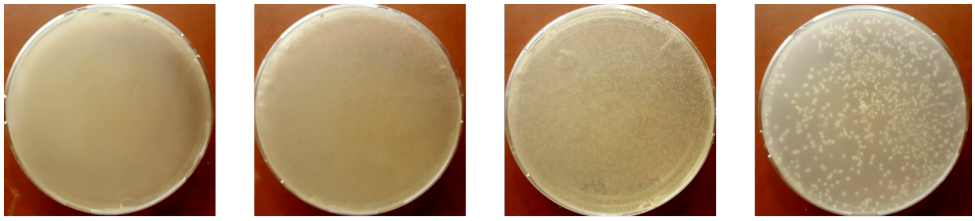


Figure S21. Optimization of the plating conditions. *E. coli* plated on agar plates (25 mL LB-agar, 10 cm diameter). OD_{600} = 1 × 10^{-6} was the optimal concentration of the cells, plated using either glass beads or sterile loops.

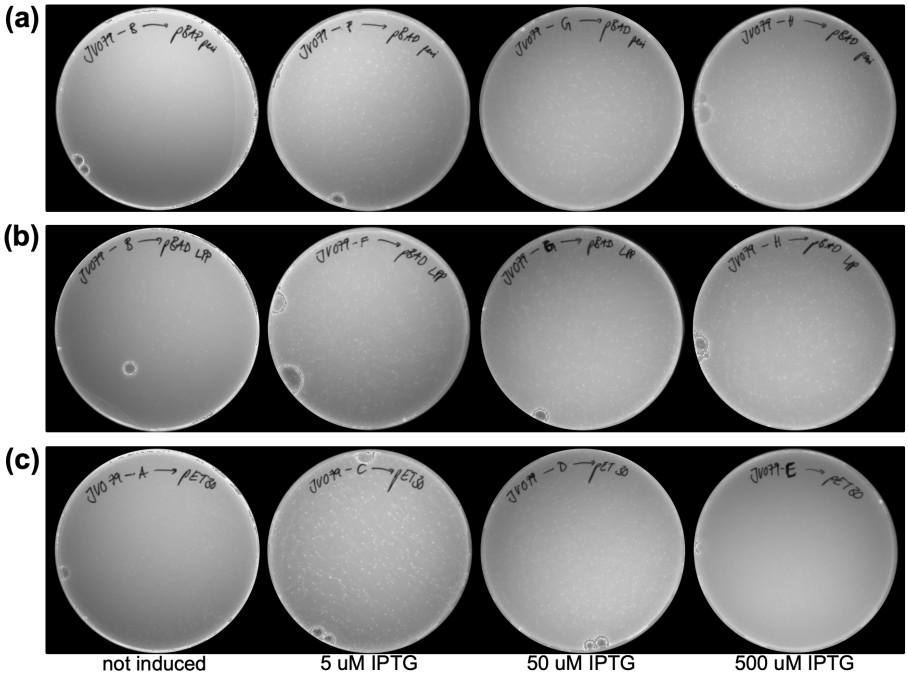


Figure S22. Expression conditions affecting the growth of *E. coli* expressing Sav. a) Periplasmic expression of Sav in the pBAD33 vector. **b)** Surface-displayed Sav in the pBAD33 vector. **c)** Periplasmic expression of Sav in the pET30b vector. Periplasmic expression of Sav using the pET30b vector and induction with low amounts of IPTG (5 µM) gave the most viable colonies. Too strong induction (50 and 500 µM) was detrimental for cell growth in the observed period of 24 h.

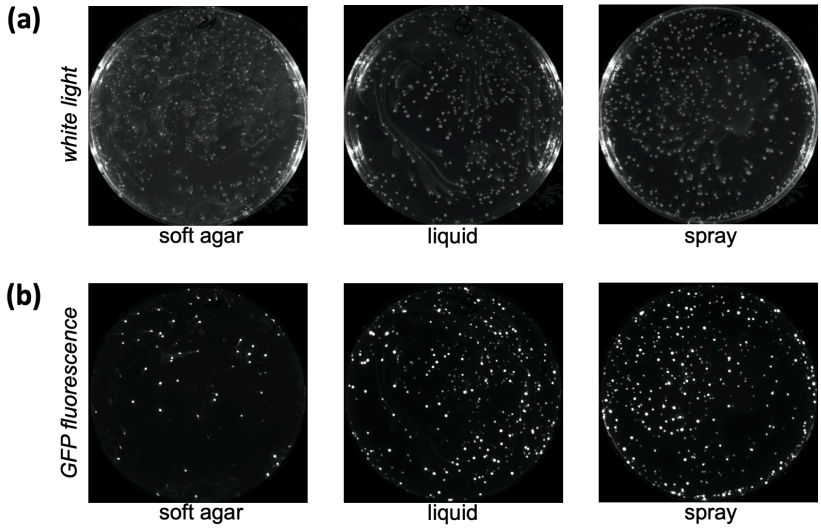


Figure S23. Test expression of GFP encoded on a pDmpR vector using aminophenol 75 as inducer. The inducer was added after growth using different techniques. Spraying the inducer onto the plate resulted in the least number of smears.

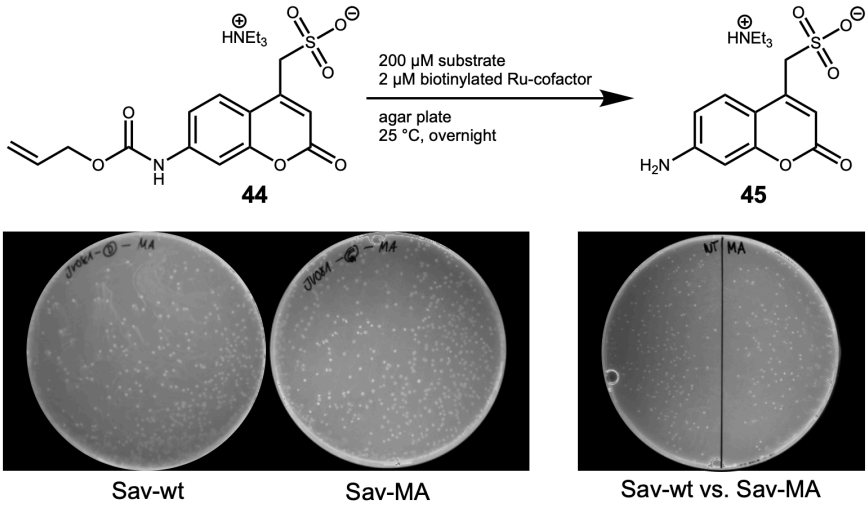


Figure S24. Agar plate screening of Sav-wt *vs.* Sav-MA in a 1:1 ratio for the deprotection of the protected aminocoumarin **44**.

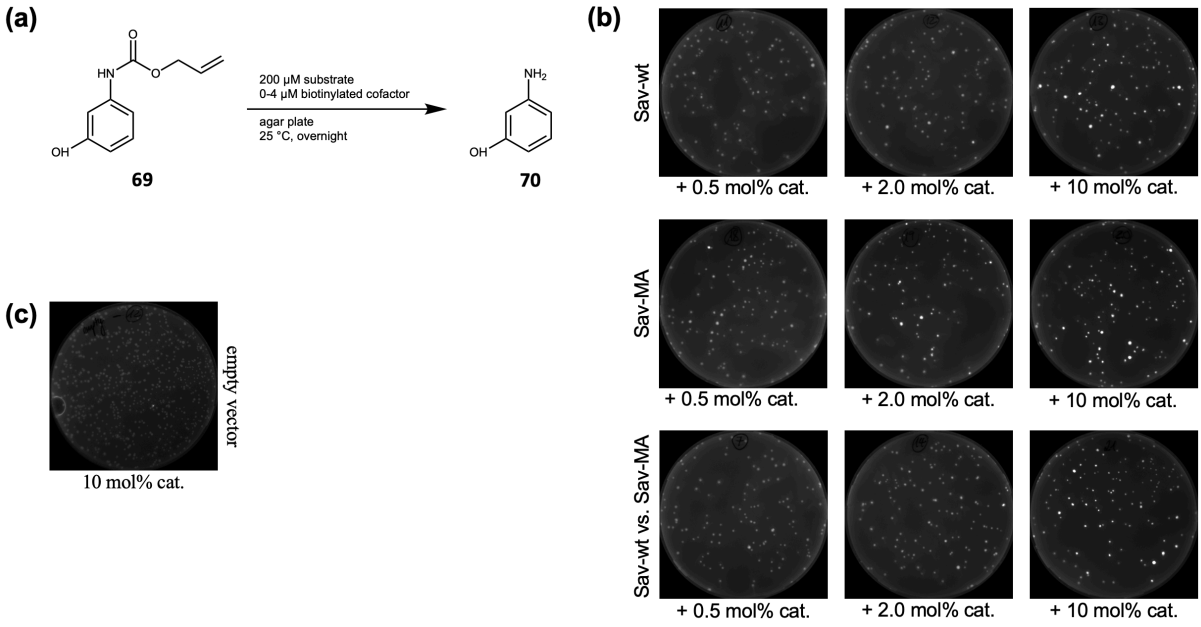


Figure S25. Agar plate screening of the deprotection of 3-aminophenol 74. **a)** Deprotection of aminophenol which turns on the overexpression of GFP. **b)** Fluorescence imaging of Sav-wt, Sav-MA and Sav-wt *vs.* Sav-MA in a 1:1 ratio. **c)** negative control = empty vector with substrate **74** and the highest concentration of cofactor **46**.

C1.2 Other Reactions and Scaffolds tested for DE Screening

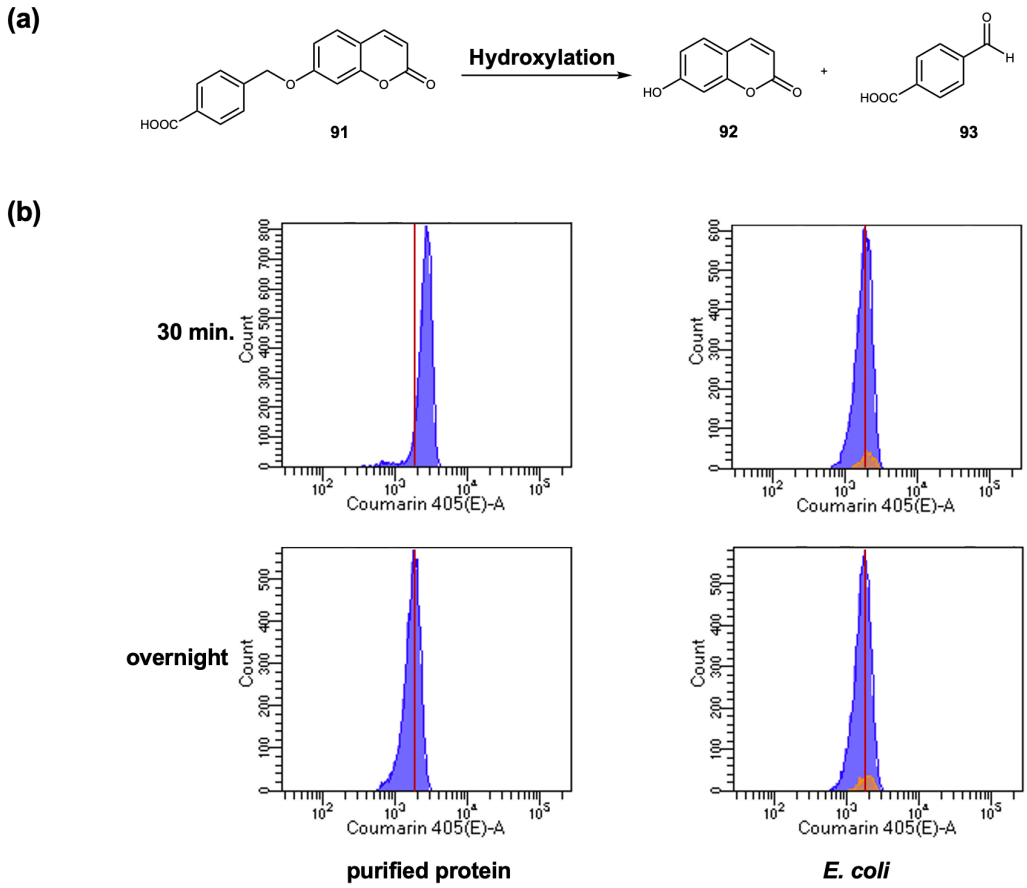


Figure S26. Hydroxylation. a) Fe-TAML · Sav catalyzed hydroxylation of the caged coumarin **91** forms coumarin **92** and **93**.[180] **b)** The reaction was encapsulated in droplets together with purified protein (left) and *E. coli* cells expressing Sav (right). The background signal is shown in blue and the *E. coli* population is shown in orange. Over time, the signals decrease in fluorescence intensities, which is most probably attributed to the poor retention of the substrate and product molecule within the water droplet.

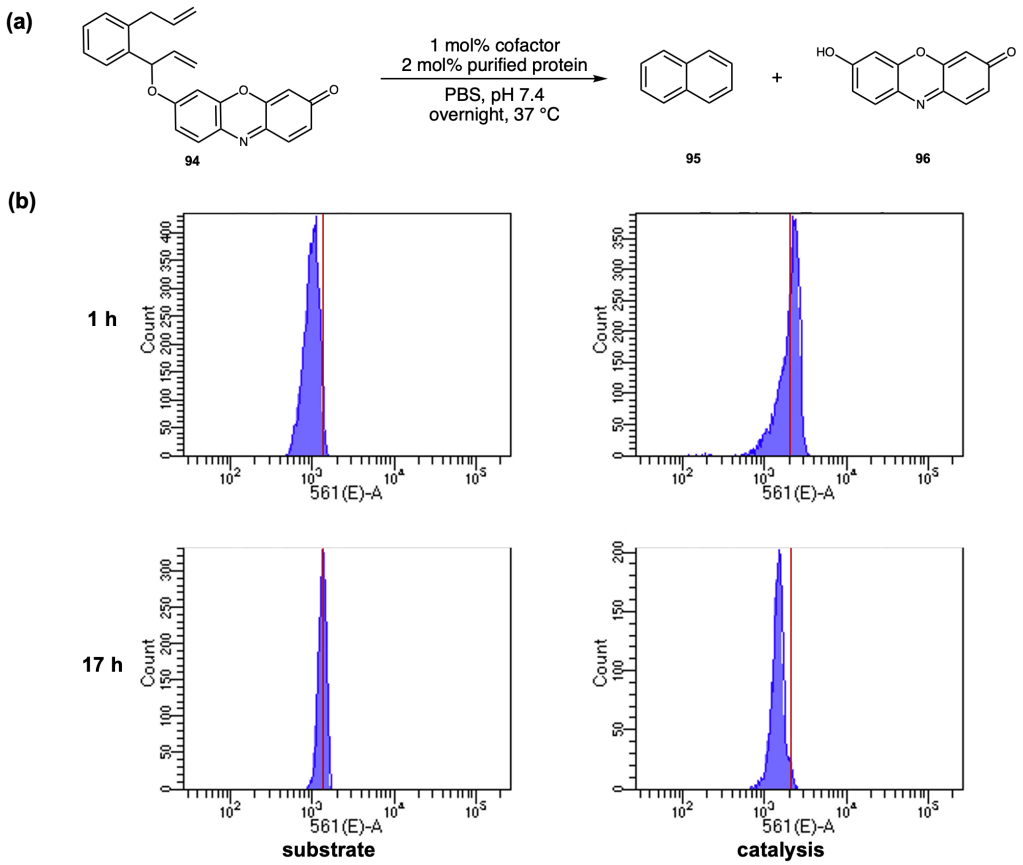


Figure S27. Metathesis. a) Ring closing metathesis leads to the formation of naphthalene **95** and resorufin **96**. **b)** The substrate was separately encapsulated (left) and showed a slight increase in fluorescence intensity overnight. Purified protein together with the substrate and the cofactor were also encapsulated and here again a decrease in fluorescence intensities was observed.

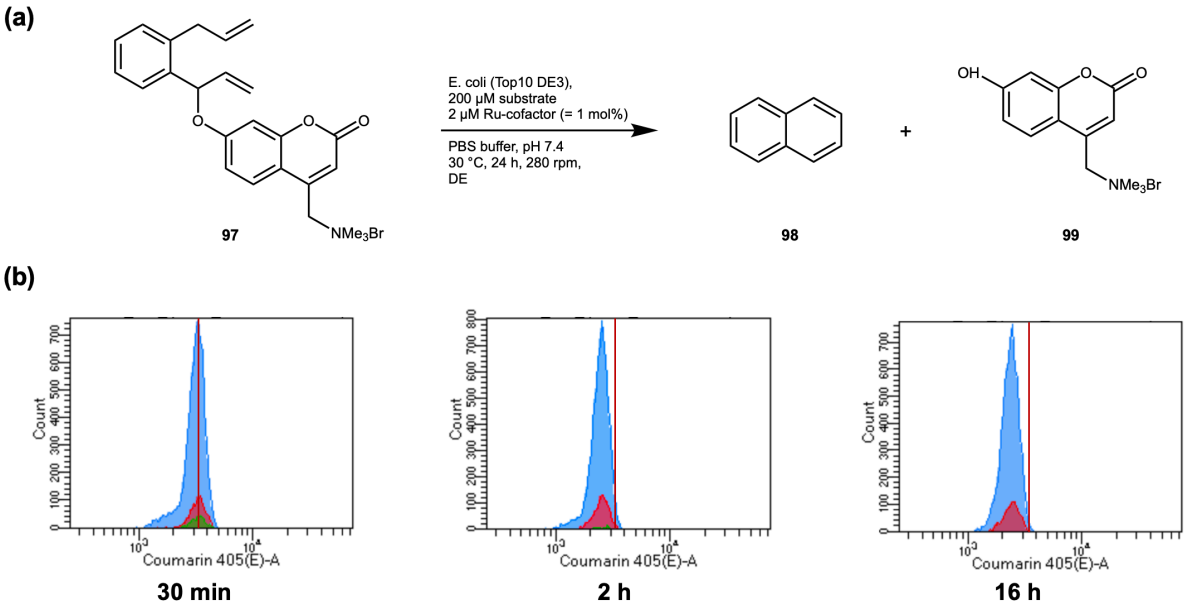


Figure S28. Metathesis. a) Ring closing metathesis leads to the formation of naphthalene **98** and a charged coumarin molecule **99**. **b)** *E. coli* expressing Sav was encapsulated together with the cofactor and the substrate. The background signal is shown in blue and the *E. coli* population is shown in red. After 2 h, the signal decreased in fluorescence intensity but did not further increase overnight.

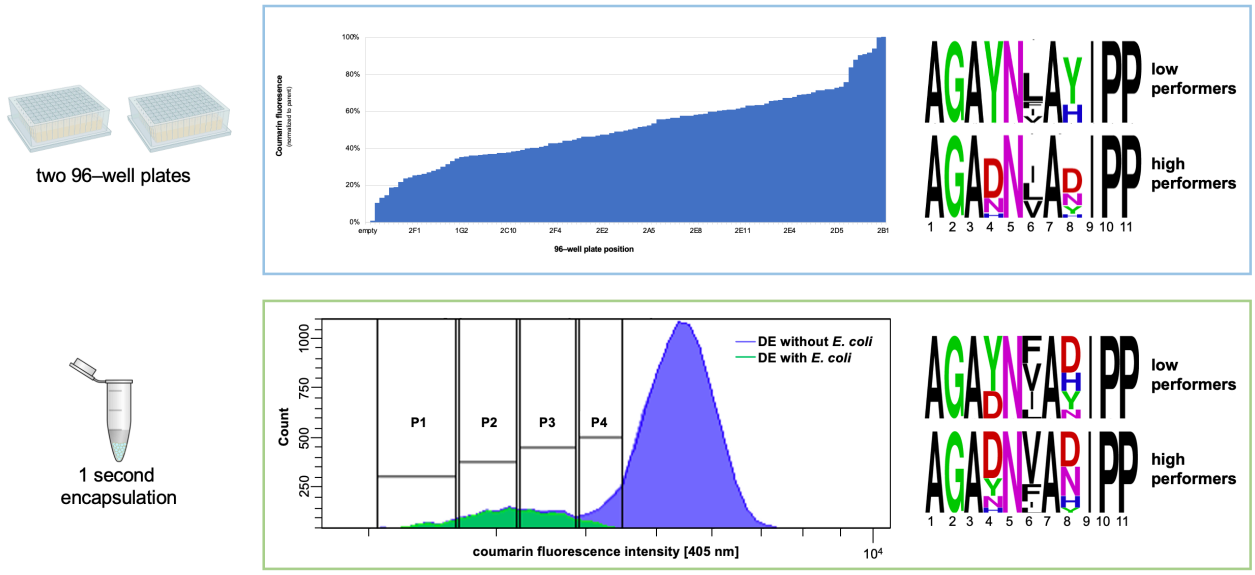


Figure S29. Initial library screening of surface-displayed Sav-SOD for the deprotection of protected aminocoumarin 44. A library of 64 variants (four different mutants at positions 4, 6 and 8 of the SOD) was screened in 96-well plates (blue) and in droplets (green). Comparison of the high performers (highest 5%) and low performers (lowest 10%) between the two screenings revealed similarities. Among the low performers tyrosine (Y) was very prominent in the 96-well plate screening and was also observed in the DE screening. Among the high performers, aspartate (D) was prominent at positions 4 and 8 in the 96-well plate screening and was also observed in the DE screening. However, this scaffold was not further investigated because of the generally low activity and the later noticed issues with surface-display of Sav.

C1.3 Droplet Assay Development

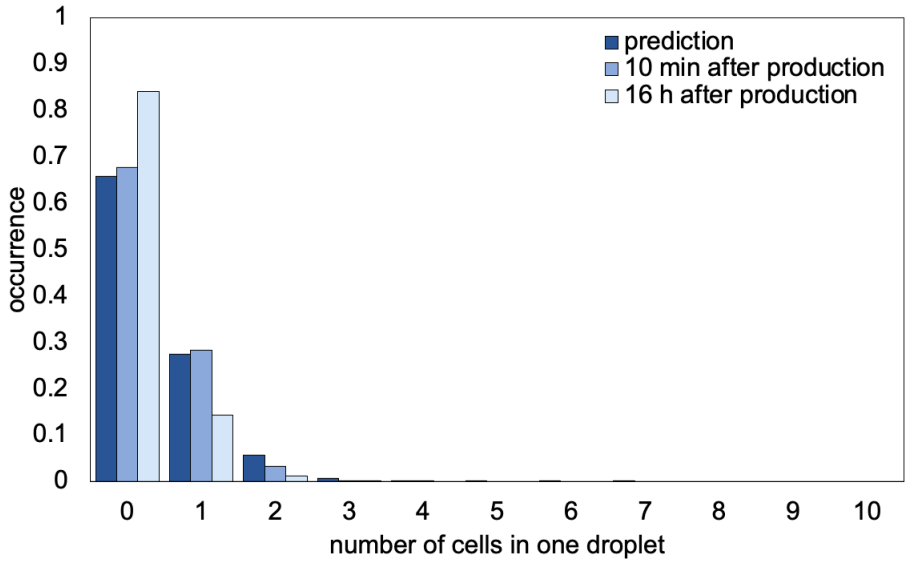


Figure S30. Poisson distribution of encapsulated single *E. coli* cells expressing Sav. After expression of Sav, the cells were incubated with the biotinylated fluorescent dye Atto-565 prior to encapsulation in double emulsion droplets. The distribution was determined by trapping the double emulsions on-chip using a trapping chip and counting the presence of an *E. coli* cell with Atto-565 fluorescence intensity.

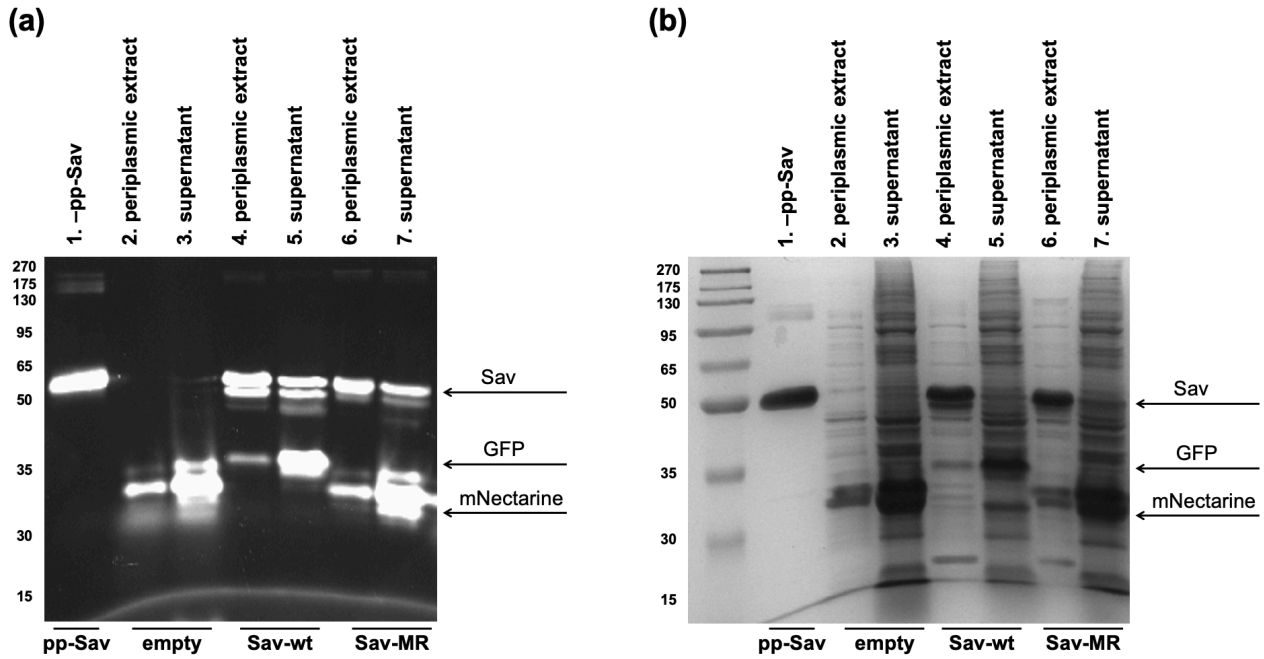


Figure S31. Expression levels of periplasmic expressed Sav-wt and Sav-MR. Extraction of the periplasmic part was performed as described elsewhere.[328] Sav-wt was labeled with GFP and Sav-MR was labeled with mNectarine. The presence of Sav in the periplasm (lane 4/6), as well as the markers GFP (lane 4/5) and mNectarine (lane 2,3,6,7) was confirmed. Further lysis of the cells revelead some amounts of Sav in the cytoplasm (lane 5/7). The markers were mainly expressed in the cytoplasm (5/7).

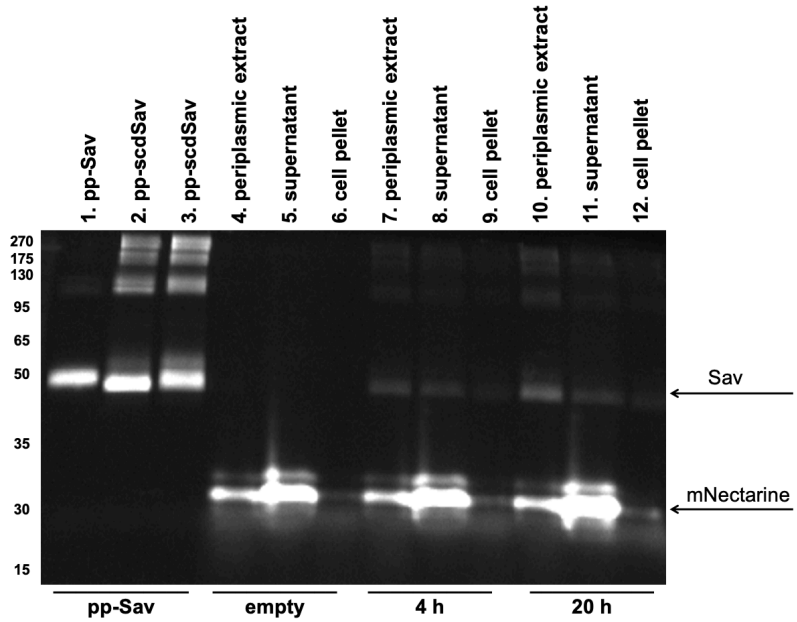


Figure S32. Expression levels of periplasmic expressed scdSav labeled with mNectarine. Extraction of the periplasmic part was performed as described elsewhere. Two different expression conditions were applied: 4 h at 25 °C (lane 7-9), and 20 h at 20 °C (lane 10-12). Overnight expression of Sav afforded slightly more protein but the expression levels were lower compared to Sav.

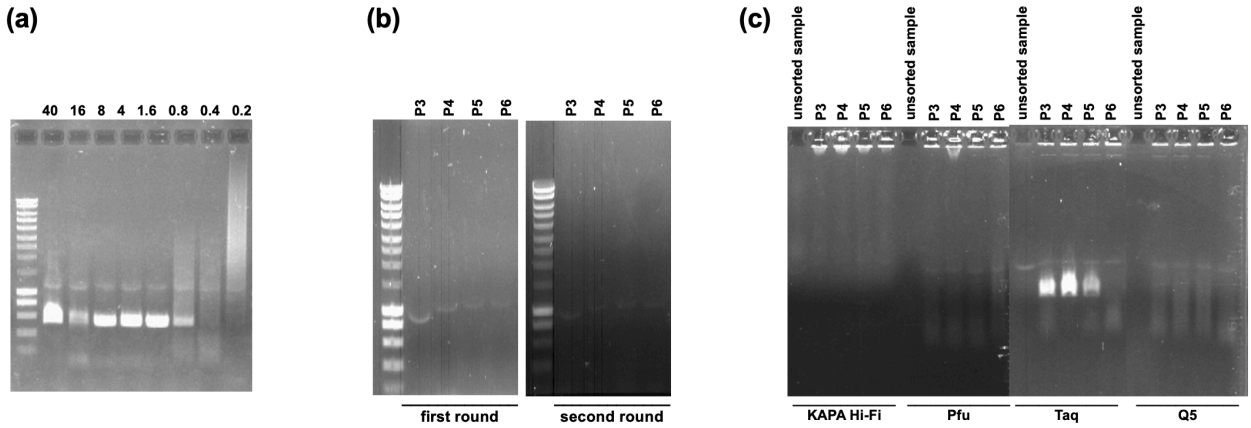


Figure S33. Optimization of PCR conditions. a) Dilution series (0.2-40 pg/µL) of parent plasmid amplified with Q5 High fidelity polymerase. **b)** Nested PCR on sorted samples. **c)** Comparison of polymerases KAPA-Hi-Fi, Pfu, Taq and Q5: Taq polymerase was the most efficient and was used for all further experiments. Moreover, Taq polymerase was most suitable because of the 3'-Alanine overhang, that is required for TOPO cloning and adapter ligation for Nanopore sequencing.

Table S8. Extended data of control sorts using Sanger sequencing, Nanopore Sequencing and NGS.

entry	sample	percentage	
		Sav-wt [50%]	Sav-MR [50%]
1[a]	unsorted (1:1)	50	50
2[a]	highest 5% (1:1)	10	90
3[a]	unsorted (9:1)	62	38
4[a]	highest 5% (9:1)	48	52
5[b]	unsorted (1:1)	69	31
6[b]	highest 5% (1:1)	45	55
7[b]	unsorted (9:1)	95	5
8[b]	highest 5% (9:1)	86	14
9[c]	unsorted (1:1)	49	51
10[c]	highest 5% (1:1)	21	76
11[c]	unsorted (99:1)	98	2
12[c]	highest 5% (99:1)	53	47

a) obtained by TOPO cloning followed by Sanger sequencing. b) obtained by Nanopore sequencing. c) obtained by NGS.

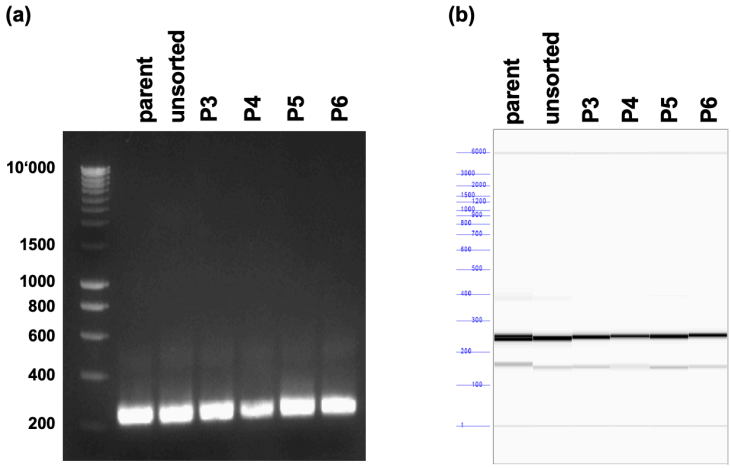


Figure S34. PCR amplification of the sort conducted on the Sav 400-variant library. An unsorted sample as well as the four sorted gates (P3-P6) were amplified using the NGS primers (Table S19). Expected size after PCR amplification is ~240 bp. The main band can be found at 240 bp. **a)** Agarose gel electrophoresis of the PCR amplification shows a good PCR amplification of the different gates. The fragments were extracted from the gel and purified according to manufacturer's protocol (Monarch gel extraction kit, NEB). **b)** Purified fragments analyzed by capillary

electrophoresis. The main band appears at ~240 bp. Slight impurities including smaller sized fragments were visible. The sample was used for Illumina sequencing without further purification. Because of the high number of reads the impurities were not an issue for data analysis.

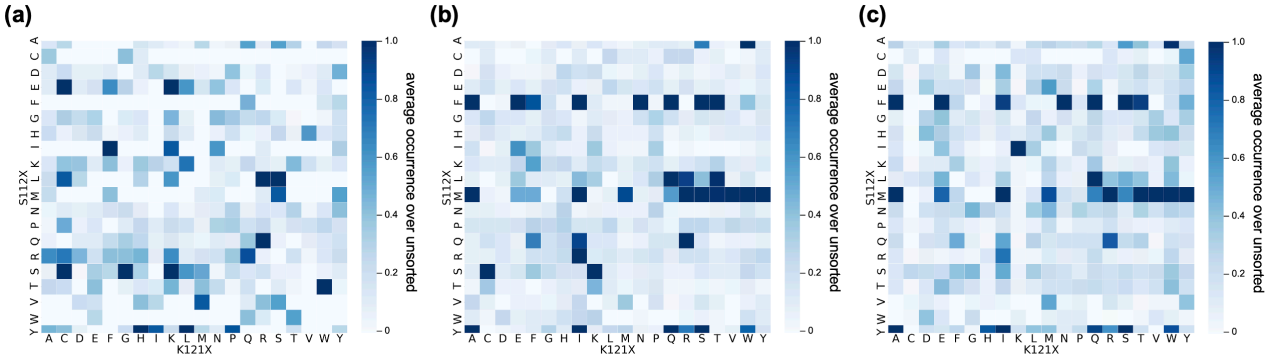


Figure S35. Extended data for the screening of the 400-variants library in DEs. a) Enrichment of the 400 mutants over the unsorted sample in the lowest gate (P3). **b)** Enrichment of the 400 mutants over the unsorted sample in the gate P4. **c)** Enrichment of the 400 mutants over the unsorted sample in the second highest gate (P5). **a-c)** x-axis: amino acids at position K121, y-axis: amino acids at position S112, displayed are relative occurrences of the respective mutants in relation to the overall reads in the whole sample.

Table S9. Homology analysis of monomer A *vs.* monomer B of scdSav.

name	sequence
monomer_A	gcaagcatgactggaggtcagcaaatgggtCGTGATCAGGCAGGTATTACCGGCACCTGG
monomer_B	GCTAGCATGACTGGTGGACAGCAAATGGGTCGGGATCAGGCCGGCATAACCGGCACCTGG
	** *********** ** ************** ******** ** ** ************
monomer_A	TATAATCAaCTGGGTAGCACCTTTATTGTTACTGCAGGCGCAGATGGTGCACTGACCGGT
monomer_B	TACGCCCAGCTCGGCGATACCTTCATCGTGACCGCGGGCGCCGACGGCGCCCTGACCGGA
	** ** ** ** ***** ** ** ** ** ***** ** ** ** ********
monomer_A	ACGTATGAAAGCGCAGTTGGTAATGCAGAAAGCCGTTATGTTCTGACAGGTCGTTATGAT
monomer_B	ACCTACGTCACGGCGCGTGGCAACGCCGAGAGCAGATACGTCCTGACCGGTCGTTACGAC
	** ** * * ** *** ** ** ** *** * ** ** ***** ******** **
monomer_A	AGCGCACCGGCAACCGATGGTAGCGGCACCGCACTGGGTTGGACCGTTGCATGGAAAAAT
monomer_B	AGCGCCCCAGCCACCGACGGCTCTGGCACCGCCCTCGGTTGGACGGTGGCCTGGAAGAAC
	***** ** ** ***** ** ******** ** ******** ** ** ***** **
monomer_A	AACTATCGTAATGCACATAGCGCAACCACCTGGTCAGGTCAGTATGTTGGTGGTGCAGAA
monomer_B	AATTACAGAAACGCCCACTCCGCGACCACGTGGAGCGGCCAATACGTCGGCGGCGCCGAG
	** ** * ** ** ** *** ***** *** ** ** ** ** ** ** ** **
monomer_A	GCACGCATTAATACCCAGTGGCTGCTGACCAGCGGCACCACCGAAGCAAATGCCTGGGCA
monomer_B	GCGAGGATCAACACACAATGGTTATTAACAAAAGGAACTACTGAGGCCAACGCATGGAAG
	** * ** ** ** ** *** * * ** * ** ** ** ** ** ** ***
monomer_A	AGCACCCTGGTTGGTCATGATACCTTTACCAAAGTTAAACCGAGCGCAGCAAGCATTGAT
monomer_B	TCCACGCTGGTCGGCTGCGCCACCTTCACCAAGGTGAAGCCTTCCGCCGCCTCAATCGAC
	*** ***** ** * ***** ***** ** ** ** *** ** ** **
monomer_A	GCAGCAAAAAAAGCCGGTGTGAATAATGGTAATCCGCTGGATGCAGTTCAGCAG---
monomer_B	GCGGCGAAGAAGGCTGGCGTCAACAACGGCAACCCTCTCGACGCCGTACAACAATAA
	** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** **

turquoise = region where the sequence skips during PCR (see Table S10).
yellow = region where the sequence skips during NGS (see Table S11).
red = positions S112 and K121.

Table S10. Details of the regions skipped during PCR amplification.

name	sequence
monomer A	TGAAAGCGCAGTTGGTAATGCAGAAAGCCGTTATGTTCTGACAGGTCGTTATGATAGCGC
monomer B	CGTCACGGCGCGTGGCAACGCCGAGAGCAGATACGTCCTGACCGGTCGTTACGACAGCGC
	* * ** *** ** ** ** *** * ** ** ***** ******** ** *****
monomer A	ACCGGCAACCGATGGTAGCGGCACCGCACTGGGTTGGACCGTTGCATGGAAAAATAACTA
monomer B	CCCAGCCACCGACGGCTCTGGCACCGCCCTCGGTTGGACGGTGGCCTGGAAGAACAATTA
	** ** ***** ** ******** ** ******** ** ** ***** ** ** **

turquoise, red, yellow, green = identified skipped regions

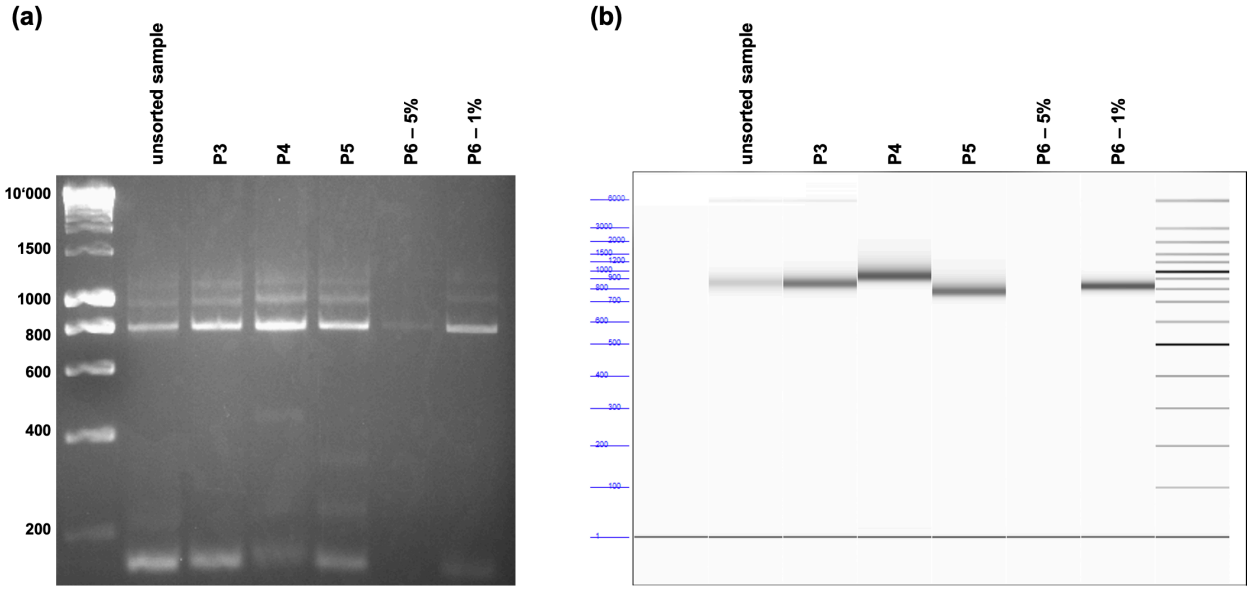


Figure S36. PCR amplification of the sort conducted on the scdSav 160'000-variant library. Unsorted sample and the sorted gates (P3-P6) were amplified using the primers XX and XX. Expected size after PCR amplification is ~800 bp. The main band can be found at 800 bp, however multiple larger and smaller fragment sizes are observed. **a)** Agarose gel electrophoresis of the PCR amplification. **b)** The main band at 800 bp was cut out of the gel, purified, and analyzed by capillary electrophoresis. This highly pure sample was used for Illumina sequencing.

Table S11. Details of the regions skipped during NGS.

name	sequence
monomer A	GGTCAGTATGTTGGTGGTGCAGAAGCACGCATTAATACCCAGTGGCTGCTGACC
monomer B	GGCCAATACGTCGGCGGCGCCGAGGCGAGGATCAACACACAATGGTTATTAACA
	** ** ** ** ** ** ** ** ** * ** ** ** ** *** * * **
monomer A	GGCCAATACGTCGGCGGCGCCGAGGCGAGGATCAACACACAATGGTTATTAACA
monomer B	AAAGGAACTACTGAGGCCAACGCATGGAAGTCCACGCTGGTCGGCTGCGCCACC
	* ** ** ** ** ** ** ** *** *** ***** ** * ***
monomer A	TTTACCAAAGTTAAACCGAGCGCAGCAAGCATTGAT
monomer B	TTCACCAAGGTGAAGCCTTCCGCCGCCTCAATCGAC
	** ***** ** ** ** *** ** ** **

red = positions S112 and K121
orange = primer binding region of NGS primers
turquoise, green = identified skipped regions

Table S12. Summarized Nanopore sequencing results for the screening of scdSav in double emulsions.

entry	S112^{A}	K121^{A}	S112^{B}	K121^{B}	reads	entry	S112^{A}	K121^{B}	S112^{A}	K121^{B}	reads
1	Thr	Gly	Met	Ser	17	37	Glu	Gly	Gln	Pro	1
2	Trp	Tyr	Lys	Lys	15	38	Glu	Ser	Thr	Asn	1
3	Leu	Met	Met	Ser	8	39	Glu	Ser	Asn	Asn	1
4	Lys	Arg	Ser	Tyr	4	40	Gly	Asp	Arg	Gln	1
5	Ala	Gln	Asn	Trp	4	41	Gly	Met	His	Asn	1
6	Gln	Phe	Cys	Phe	4	42	Gly	Ile	Asn	Trp	1
7	Ala	Gly	Val	Gly	3	43	His	Gly	Met	Ser	1
8	Lys	Lys	Val	Met	3	44	Ile	Glu	Trp	Cys	1
9	Val	Ile	Gln	Leu	3	45	Ile	Gly	Cys	Gly	1
10	Ala	Gly	Val	Gly	2	46	Ile	Met	Met	Ser	1
11	Ser	Met	Met	Ser	2	47	Leu	Stp	Met	Ser	1
12	Thr	Glu	Met	Ser	2	48	Lys	Glu	Met	Gln	1
13	Thr	Ile	Phe	Lys	2	49	Lys	Lys	Val	Ile	1
14	Ala	Gln	Asp	Trp	1	50	Met	Asn	His	Arg	1
15	Ala	Gly	Arg	Gln	1	51	Phe	Ile	Thr	Leu	1
16	Ala	Lys	Val	Met	1	52	Pro	Tyr	Lys	Lys	1
17	Ala	Arg	Gln	Phe	1	53	Pro	Arg	Ser	Stp	1
18	Arg	Ile	Phe	Lys	1	54	Ser	Lys	Met	Ser	1
19	Arg	Ile	Phe	Asn	1	55	Stp	Gly	Thr	Ser	1
20	Arg	Met	His	Asn	1	56	Thr	Gly	Glu	Ser	1
21	Arg	His	Arg	Phe	1	57	Thr	Gly	Asn	Ser	1
22	Arg	Lys	Val	Met	1	58	Thr	Gly	Leu	Ser	1
23	Asn	His	Arg	Arg	1	59	Thr	Gly	Pro	Ser	1
24	Asn	His	Stp	Arg	1	60	Thr	Ile	His	Ser	1
25	Asn	His	His	Arg	1	61	Thr	Thr	Cys	Asp	1
26	Asp	Gln	Phe	Val	1	62	Thr	Thr	Val	Met	1
27	Asp	Ile	Phe	Lys	1	63	Thr	Thr	Val	Ser	1
28	Asp	Gln	Leu	Thr	1	64	Thr	Gly	His	Ser	1
29	Gln	Asn	Ala	Trp	1	65	Thr	Arg	Val	Ser	1
30	Gln	Gly	Met	Ser	1	66	Thr	Asn	Val	Ser	1
31	Gln	Phe	Met	Phe	1	67	Thr	Gly	Thr	Ser	1
32	Gln	Phe	Cys	Phe	1	68	Trp	Phe	Cys	Phe	1
33	Gln	Ser	Ser	Ala	1	69	Trp	Tyr	Ser	Lys	1
34	Gln	Tyr	Ala	Stp	1	70	Trp	Tyr	Ser	Tyr	1
35	Gln	Tyr	Lys	Lys	1	71	Trp	Tyr	Cys	Lys	1
36	Glu	Arg	Ser	Tyr	1	72	Trp	Tyr	Leu	Lys	1

total reads obtained by Nanopore sequencing: 526, correctly aligned reads: 128.
Different colors depict the occurrence of the same pattern in the second monomer across all aligned reads.

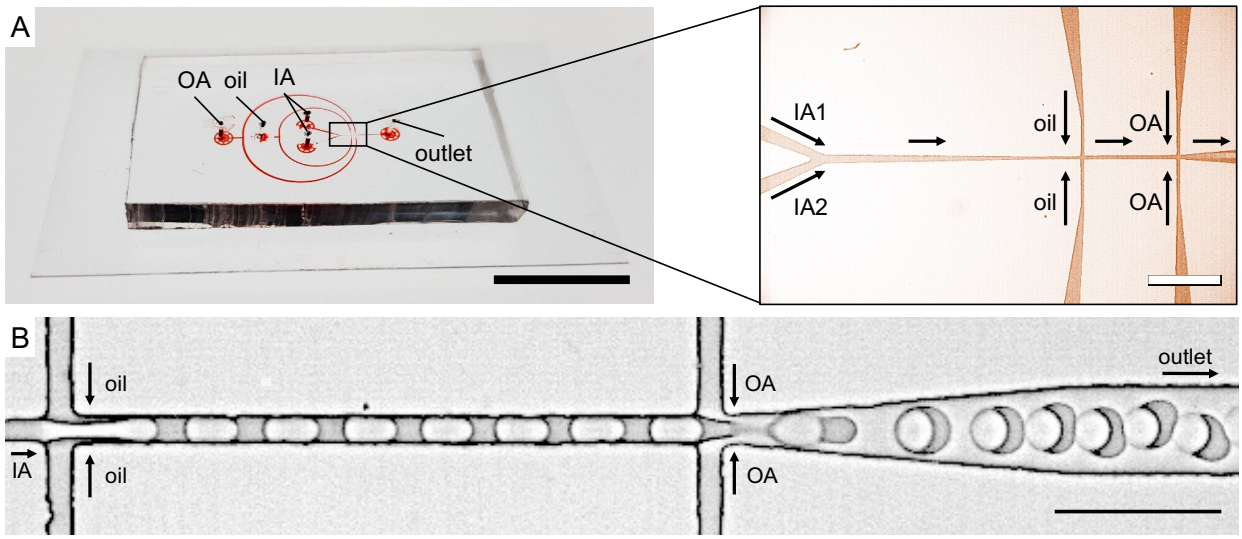


Figure S37. Microfluidic chip for the double emulsion production. a) Photograph and micrograph of the microfluidic device designed for DE formation. The device has channels with a height of 11 µm and contains 4 inlets: one inlet for the outer aqueous phase (OA), one inlet for the oil (oil) and two inlets for the inner aqueous phases (IA). Black scale bar: 1 cm. White scale bar: 200 µm. **b)** Micrograph of DE formation. Monodisperse DEs co-encapsulating *E. coli* cells, cofactor **1** and substrate **2** are produced using a PDMS-based microfluidic device. Scale bar: 50 µm.

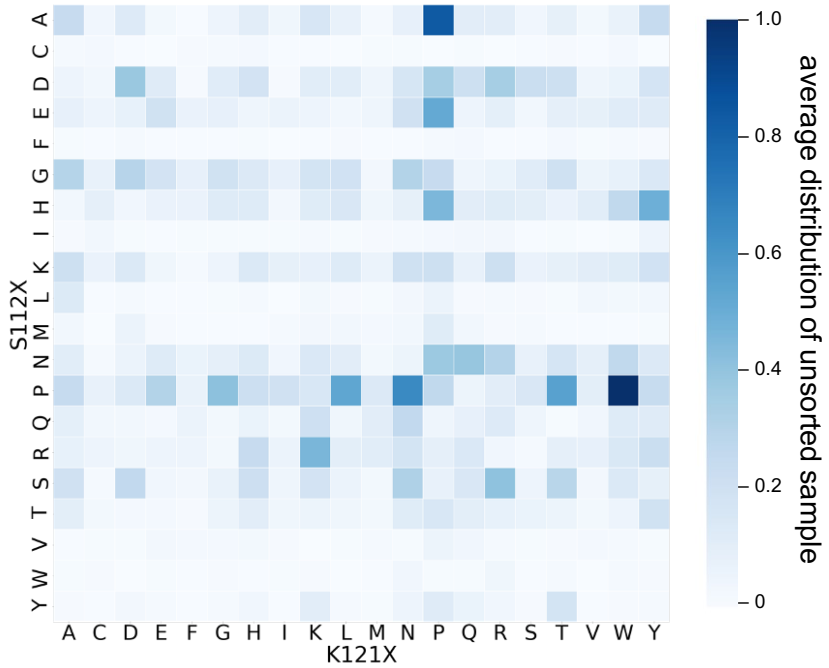
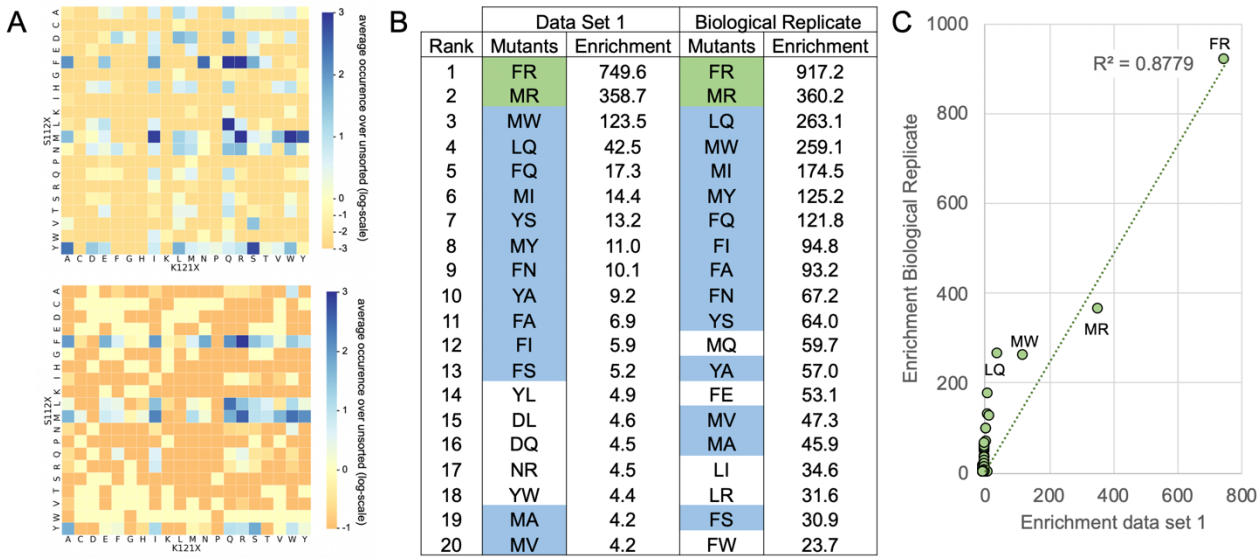


Figure S38. Screening of a 400-variant library in DEs. Relative distribution of the 400 mutants in the unsorted sample determined by NGS. *x* axis: amino acids at position K121, *y* axis: amino acids at position S112. Displayed are relative occurrences of the respective mutants in relation to the overall reads in the whole sample determined by NGS.



	Data Set 1		Biological Replicate	
Rank	Mutants	Enrichment	Mutants	Enrichment
1	FR	749.6	FR	917.2
2	MR	358.7	MR	360.2
3	MW	123.5	LQ	263.1
4	LQ	42.5	MW	259.1
5	FQ	17.3	MI	174.5
6	MI	14.4	MY	125.2
7	YS	13.2	FQ	121.8
8	MY	11.0	FI	94.8
9	FN	10.1	FA	93.2
10	YA	9.2	FN	67.2
11	FA	6.9	YS	64.0
12	FI	5.9	MQ	59.7
13	FS	5.2	YA	57.0
14	YL	4.9	FE	53.1
15	DL	4.6	MV	47.3
16	DQ	4.5	MA	45.9
17	NR	4.5	LI	34.6
18	YW	4.4	LR	31.6
19	MA	4.2	FS	30.9
20	MV	4.2	FW	23.7

Figure S39. Screening validation and data reproducibility for the 400-variant library. a) Heat map comparison between enrichments found for data set 1 (top) and biological replicate (bottom). **b)** Enrichment values for the top 20 mutants in data set 1 and its biological replicate. Green: mutants found in the top 20 of both screenings at the same rank. Blue: mutants found in the top 20 of both screenings. **c)** Correlation curve between the enrichment values found for data set 1 and its biological replicate.

Table S13. Enrichment factor for the top 20 hits of the DE screening. The enrichment was computed as (count mutantX in top 5%)/(total count in top 5%)×(total count in DE)/(count mutantX in DE).

Mutant	DE count	Top 5% count	Enrichment in top 5%
FR	416	311845	741.8
MR	1072	384548	355.0
MW	575	71024	122.2
LQ	1201	51067	42.1
FQ	1642	28326	17.1
MI	1441	20822	14.3
YS	1022	13524	13.1
MY	834	9160	10.9
FN	325	3271	10.0
YA	1176	10761	9.1
FA	749	5136	6.8
FI	153	901	5.8
FS	601	3106	5.1
YL	1395	6892	4.9
DL	8849	40409	4.5
DQ	16938	76273	4.5
NR	24176	108007	4.4
YW	1037	4546	4.3
MA	2569	10807	4.2
MV	503	2093	4.1

Table S14. Comparison of the top 20 mutants between 96-well plate experiment (activity measurements) and DE screening (enrichment). Overall, we found a 50% match in the top 10 hits and a 60% match in the top 20 hits.

rank	96-well plate screening	Activity (improvement over wt-Sav)	DE screening	Enrichment (over unsorted sample)
1	FQ	13.8	FR	741.8
2	FR	13.4	MR	355.0
3	MR	13.4	MW	122.2
4	MW	13.3	LQ	42.1
5	MI	13.0	FQ	17.1
6	AW	12.0	MI	14.3
7	FS	11.1	YS	13.1
8	MM	10.9	MY	10.9
9	LR	10.8	FN	10.0
10	FT	10.5	YA	9.1
11	MV	10.3	FA	6.8
12	FA	10.3	FI	5.8
13	FN	10.2	FS	5.1
14	FE	10.0	YL	4.9
15	YS	9.7	DL	4.5
16	IR	9.5	DQ	4.5
17	QR	9.4	NR	4.4
18	QI	9.2	YW	4.3
19	MA	9.1	MA	4.2
20	MY	9.1	MV	4.1

C.2 Experimental Part

C.2.1 Materials

If not otherwise stated, PBS was bought from Bioconcept. 1 H, 1 H, 2 H, 2H-perfluorooctanol was bought from Fluorochem. Tris-(acetonitril)-cyclopentadienylruthenium(II)-hexafluoro-phosphate was purchased from Sigma Aldrich. Water was purified with a Milli-Q-system (Millipore). Antibiotics were purchased from Applichem GmbH. All enzymes, the Monarch plasmid extraction kit, the Monarch PCR and DNA clean-up kit, and the Gibson assembly master mix, were purchased from New England BioLabs. Magnetic beads used for DNA purification were AMPure XP purchased from Beckman-Coulter Life Sciences. Sodium dodecyl sulfate (SDS) was bought from abcr. Hydrofluoroether Poly(dimethylsiloxane) (PDMS, Sylgard 184) was purchased from Dow Corning. (HFE 7500) with 5% 008-FluoroSurfactant was purchased from Ran biotechnologies. All supplies for Nanopore sequencing were purchased from Oxford Nanopore Technologies.

The bacterial strain used is Top10 (DE3) (Invitrogen: F^- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*araleu*)7697 *gal*U *gal*K *rps*L (Str^R) *end*A1 *nup*G). Chemically competent cells (prepared according to the RbCl-method following the Hanahan protocol) and electrocompetent cells bearing the mNectarine plasmid were prepared in-lab.

The pET30b vector was used for the generation of the mutant library. The pSC101 was used for mNectarine and was generously provided by the laboratory of Prof. Jeschek/Prof. Panke. Primers for cloning and NGS were designed individually and were synthesized at IDT technologies or Microsynth AG. The 1.6×10^5 variant library was bought as a combinatorial variant library at Twist Bioscience.

C.2.2 Methods

For all biological experiments, the equipment was sterilized (121 °C, 20 min). PCR reactions were performed with an Eppendorf Mastercycler Gradient following a general QuikChange™ protocol (Table S16). Agarose and SDS gel electrophoresis chambers and the MicroPulsor Electroporator were purchased from Bio-Rad Laboratories Inc. The gels were visualized with the software Quantity One. DNA concentrations were measured with Nanodrop1000 or the Advanced Analytical 12-capillary Fragment Analyzer from Agilent. Fluorescence scans of individual substances were conducted using a Tecan Infinite M1000 Pro. Large scale protein purification was done using ÄKTA prime Fast Liquid Protein Chromatography System (GE Healthcare). Double

emulsion were analyzed on a BD LSR Fortessa SORP (Special Order Research Product) and sorted on a BD FACSAria™ II cell sorter. NGS was carried out on an Illumina NextSeq 500 platform.

For the microfluidics assay, neMESYS syringe pumps (Cetoni) and 1 mL syringes (VWR, BD Plastipak Luer-lock) were used to introduce the solutions into the microfluidic chip. We used PTFE tubing (ID = 0.56 mm, Adtech Polymer Engineering™), precision dispenser needles (23 gauge, Metcal) and metal pins (New England Small Tube, NE-1310-03, 0.025″ OD x .013″ ID x 1.00″ length) to connect the syringes to the chip. To control the chip coating with polyvinyl alcohol (PVA), we used a MFCS-8C pressure control unit (Fluigent).

C.2.3 BUFFERS AND STOCK SOLUTIONS

Biotinylated ligand	2 mM in DMF
$[CpRu(MeCN)_3]PF_6$	2 mM in DMF
Ruthenium cofactor **46**	1 mM in DMF (10 µL of ligand stock was mixed with 10 µL $[CpRu(MeCN)_3]PF_6$ stock and incubated for 5 min)
Sav isoform	0.4 mM free biotin-binding sites in PBS-buffer
Substrate	10 mM substrate **44** in DMF
Product	20 mM product **45** in DMF
PBS buffer	50 mM NaH_2PO_4/Na_2HPO_4 (set to pH 7.4), 0.9 % NaCl
PBS + 0.5% SDS	9.5 mL of PBS were mixed with 0.5 mL of a 10% SDS stock solution.
Lysogeny Broth (LB)	1 % tryptone, 0.5 % yeast extract and 1 % NaCl were dissolved in distilled water and autoclaved.
LB_{agar}	1 % tryptone, 0.5 % yeast extract, 1 % NaCl and 1.5 % agar were dissolved in distilled water and autoclaved.
LB_{Soc}	2 % tryptone, 0.5 % yeast extract, 10 mM NaCl, 2.5 mM KCl were dissolved in distilled water.
B4F-solution	A stock solution of B4F (10 mM) in mQ water was prepared and diluted in BB to a final concentration of 0.4 mM B4F.
Buffer for SDS-PAGE	Tris-base (15.1 g), glycine (72 g) and SDS (5 g) were dissolved in distilled water (1 L)

C.2.4 LIBRARY PREPARATION

GENERAL SETTINGS

All mutagenesis experiments were carried out with the polymerase Q5 and as outlined below with appropriate adjustments in the annealing temperature/extension time. Once all reagents were added in appropriate volumes, the samples were placed in an Eppendorf Mastercycler Gradient and the QuikChange™ program was carried out (Table S16).[329]

Table S15. General reagents set-up for PCR

Reagents	Appropriate Amounts for one Reaction
mQ H_2O	9.5 µL
2x Q5 maser mix	12.5 µL
Plasmid template (25 ng/µL)	1.00 µL
Forward Primer (10 µM)	1.00 µL
Reverse Primer (10 µM)	1.00 µL
Final volume for one reaction	25.0 µL

Table S16. General PCR program for the library preparation

Cycle number	Denature	Anneal	Extend
1	95°C, 2 min		
2-25	95°C, 15 s	65/72 °C, 20 s	72°C, 2-5 min
27			72°C, 10 min
28		4 °C	

Table S17. Primers used for cloning.

Primer	Sequence
pRSFDuet _scdSav_fwd	ggcacgtggtataatcagctgggtagc
pRSFDuet_scdSav_rev	ttattgttgtacggcgtcgagaggg
pET30b_vector_fwd	gccgtacaacaataaggatccgaattcg
pET30b_vector_rev	gattataccacgtgccggtaatacctgc
scdSav_CVL_fwd	ggcacgtggtataatcagctgggtagc
scdSav_CVL_rev	ttattgttgtacggcgtcgagaggg
pET30b_scdSav_vector_fwd	gccgtacaacaataaggatccgaattcg
pET30b_scdSav_vector_rev	gattataccacgtgccggtaatacctgc
MGG_backbone_fwd	gatcatggtctccgacaagcttgcggccgc
MGG_backbone_rev	gatcatggtctcgtgctacccagttgattatacc
MGG_insert_fwd	gatcatggtctctagcacctttattgttactgc
MGG_insert_rev	gatcatggtctcttgtcgacggagctagaattcg

PREPARATION OF THE 400-VARIANTS LIBRARY

A glycerol stock containing *E. coli* cells with the 400-mutant library at positions S112 and K121 was grown overnight and the plasmid extracted. The purified plasmid (~100 ng) was transformed into electrocompetent Top10(DE3) cells (50 µL), containing the plasmid pSC101 with mNectarine encoded. The electroshock was applied using the MicroPulser electroporator by Bio-Rad Laboratories, Inc. Immediately after the electroshock, SOC-medium (450 µL) was added, the reaction transferred to a sterile Eppendorf tube and incubated at 37 °C for 40-60 min. The transformation was split into four equal parts and plated on big LB-agar plates supplemented with kanamycin and chloramphenicol and incubated overnight at 37 °C. All plates were scraped by the addition of LB-medium (2 mL per plate) and the combined cell suspension was aliquoted. 100 µL cell suspension was mixed with 100 µL of a glycerol stock solution (50%) to obtain 20 aliquots with

a ~25% final glycerol concentration. The glycerol stocks were immediately frozen in liquid N2 and finally stored at -80 °C. One whole such aliquot was used for the inoculation of cultures for the screening assay.

mNECTARINE PLASMID

The plasmid for the constitutive expression of mNectarine is based on pUA66[3]. The kanamycin resistance cassette was exchanged for a chloramphenicol resistance by amplifying insert and backbone with compatible overhangs and assembling them by Gibson Assembly. Likewise, GFP was exchanged for mNectarine with promoter BBa_J23111 and ribosome binding site BBa_B0030. This plasmid was generously provided by Tobias Vornholt and Markus Jeschek (ETHZ, D-BSSE).

CLONING OF THE PERIPLASMIC SCDSAV CONSTRUCT

The pET30b_scdSav was obtained by Gibson assembly of the scdSav from the pRSFDuet vector into the template vector pET30b_Sav. pRSFDuet was amplified by PCR using the primers pRSFDuet_scdSav_fwd and pRSFDuet_scdSav_rev (Table S17). The template plasmid pET30b_Sav was amplified by PCR using the primers pET30b_vector_fwd and pET30b _vector_rev (Table S17). The PCR products were purified according to manufacturer's protocol (Monarch PCR clean-up kit by NEB) and digested with DpnI overnight at 37 °C. A Gibson assembly reaction (10 µL total) was set up according to manufacturer's protocol and incubated at 50 °C for 2 h. Half of the reaction mixture was transformed directly into chemically competent cells without any further purification. The resulting pET30b_scdSav was verified by sequencing (Microsynth AG).

PREPARATION OF THE 160'000-VARIANTS LIBRARY

Four positions, S112A, K121A, S112B, and K121B were targeted for the library construction because of their close proximity to the metal cofactor and the known influence on activity. A full site saturation library was envisioned and the library was bought as a combinatorial variant library (scdSav_CVL) from Twist Biosciences. The scdSav_CVL was amplified by PCR using the primers scdSav_CVL_fwd and scdSav_CVL_rev (Table S17). The template plasmid pET30b_scdSav was amplified by PCR using the primers pET30b_scdSav_vector_fwd and pET30b_scdSav_vector_rev (Table S17). The PCR product was digested with DpnI (NEB) overnight at 37 °C. Two identical Gibson assembly reactions were set up as follows: 10 µL CVL-insert, 10 µL pET30b_scdSav and 20 µL Gibson assembly master mix (NEB). The reaction mixture was incubated over 4 h at 50 °C.

Following plasmid purification according to manufacturer's protocol (Monarch PCR clean-up kit by NEB), the plasmid (~100 ng) was transformed into electrocompetent Top10(DE3) cells (50 µL), containing the plasmid pSC101 with mNectarine encoded. The electroshock was applied using the MicroPulser electroporator by Bio-Rad Laboratories, Inc. Immediately after the electroshock, SOC-medium (450 µL) was added, the reaction was transferred to a sterile Eppendorf tube and incubated at 37 °C for 40-60 min. The 500 µL were split into two equal parts and plated on big LB-agar plates supplemented with kanamycin and chloramphenicol and incubated overnight at 37 °C. This step was repeated until the desired coverage (5.0×10^5) was reached.

The transformation efficiency was roughly 16'000 colony forming units per transformation. Therefore, the transformation was repeated 35 times in total. All plates were scraped by the addition of 2 mL LB-medium to each plate and combined. 100 µL of the combined cell suspension was mixed with 100 µL of a glycerol stock solution (50%) to obtain 80 aliquots with a ~25% (v/v) final glycerol concentration. The glycerol stocks were immediately frozen in liquid N_2 and finally stored at -80 °C. One whole such aliquot was thawed on ice and used for the inoculation of cultures for the screening assay.

C.2.4 Screening Assay

Substrate and Product Stock Solution Preparation

Ally-carbamate protected coumarin **44** and aminocoumarin **45** were synthesized as previously reported.[211,301] Stock solutions of substrate **44** (10 mM) and product **45** (10 mM) were prepared in freshly filtered PBS (0.2 µm filter), aliquoted in 60 µL samples to avoid multiple cycles of freezing and thawing and stored at -20 °C for further use.

Cofactor Preparation

The cofactor was prepared in situ by mixing a solution of $[CpRu(MeCN)_3]PF_6$ (2 mM in degassed DMF) and a solution of the biotinylated ligand (2 mM in degassed DMF) in a 1:1 ratio inside a glovebox under exclusion of oxygen. The biotinylated ligand was synthesized according to literature.[179] The cofactor stock solution (1 mM in DMF) was used outside the glovebox after incubation of 5-10 min at room temperature.

General protocol for Sav/scdSav expression for the screening

A preculture of LB medium (5 ml) supplemented with chloramphenicol (32 mg/ml) and kanamycin (50 mg/ml) was inoculated with the before prepared glycerol stock of the library of interest and incubated for 8 h at 37°C and 300 rpm. A culture of LB medium (25 ml) supplemented

with chloramphenicol (32 mg/ml) and kanamycin (50 mg/ml) in a shaking flask (250 ml) was inoculated with the preculture to a starting OD600 =0.05 and incubated for ~1-2 h at 37°C and 300 rpm (until an OD600=0.5-0.8 was reached). Sav expression was induced by the addition of IPTG (50 µM final concentration) and expression was done overnight at 25 °C and 300 rpm. One sample (1 mL) of cell culture with an OD600=0.20 was prepared and centrifuged (5 min; 17000 g), the supernatant discarded and the pellet resuspended in PBS (990 µL, pH 7.4). The cofactor stock solution (10 µL, 1 mM, 10 µM final concentration) was added to this cell suspension to afford the cells-cofactor mixture for the droplet production.

C.2.6 MICROFLUIDIC SET-UP

MICROFLUIDIC PLATFORM FABRICATION.

The devices were produced according to a protocol described previously.[303] Briefly, the microfluidic chips were produced using a SU-8 master mold on which a mixture of PDMS and curing agent (ratio 10:1) was poured. The wafer was cured at 80 °C for 3 hours. After punching inlets and outlets with a biopsy puncher (diameter 0.5 mm), the chips were plasma bonded to PDMS-coated glass slides. The device has four inlets for the outer aqueous phase (OA), the oil phase (OP) and the two inner aqueous phases (IA1 and IA2), and one outlet. To allow for DE formation, a 2.5% PVA solution was used to coat hydrophilically the OA and outlet channels according to a protocol previously described by Deshpande *et al.*[302]

MICROFLUIDIC ASSAY.

DE formation was achieved by using flow rates between 0.5 and 5 µL min^{-1} for all solutions (IA1, IA2, OA, OP) and was monitored on an inverted Olympus microscope (IX71) with a high-speed camera (Phantom VEO). In a typical experiment, the solutions contained the following components. OA: PBS with 5 g L^{-1} SDS, OP: HFE 7500 with 2% 008-FluoroSurfactant, IA1: PBS with bacteria (OD_{600}=0.2) and cofactor (10 µM) and IA2: PBS with substrate (1000 µM). The DEs were collected in a 1.5 mL Eppendorf tube *via* a short piece of PTFE tubing connected to the outlet. The Eppendorf tube was exchanged every 5 min and the encapsulation was carried out for a minimum of 40 min to ensure that the library coverage was achieved. Generally, the first Eppendorf tube was discarded to avoid non-monodisperse samples due to initial instability of the production.

DE TRAPPING AND COUNTING

The trapping was carried out on a microfluidic device as described by Stucki and Juskova *et al.*[303] Briefly, during stable droplet production, the inlet of the trapping chip was connected to the

outlet of the production chip *via* a short piece of PTFE tubing. After trapping of the DEs, the flow rate was reduced to a minimal flow (0.5 µL min^{-1}) to ensure that the DEs remain in the traps. The DEs and the *E. coli* (presence of GFP or Atto-565 signal), were counted manually using an inverted Olympus microscope (IX71). The experiment was done in triplicates and the average counts are displayed in Figure 40 and Figure S30.

FLUORESCENCE MICROSCOPY IMAGING OF DES

Fluorescence microscopy of trapped DEs was carried out using trapping arrays with microfluidic valves and as described by Stucki and Juskova *et al.*[303] Briefly, the DEs were trapped on the trapping array and imaged using an inverted microscope (NikonTi-Eclipse). The chip and the microscope were set within an environmental box and the temperature was maintained at 37 °C for the time period of catalysis. Over a period of 15 h brightfield images and fluorescence images were acquired applying the appropriate filters for detecting coumarin fluorescence. Pictures were taken ever 15 min for the first 3 h, and every 30 min for the remaining period.

DE INCUBATION AND FACS

The DEs were incubated statically in collection tubes in an oven at 37 ºC for at least 3 hours. For measurements on the flow cytometer, 0.5 µL of DE solution was added to 300 µL of OA (PBS with 5 g L^{-1} SDS), agitated manually and loaded on a BD LSR Fortessa SORP (Special Order Research Product). The DEs were gated by size on forward and side scatter profile. The DEs were then gated for singlets using bivariate plots of DEs side scatter height vs area, followed by gating for mNectarine fluorescence (λ_{ex} = 560 nm, λ_{em} = 580 nm) to determine the presence/absence of cells. Finally, the coumarin fluorescence was measured (λ_{ex} = 405 nm, λ_{em} = 450 nm). For sorting experiments, 10 µL of DE solution were added to 600 µL of OA (PBS with 5 g L-1 SDS), agitated manually and loaded on a BD FACSAria™ II cell sorter. This operation was repeated every time the sample tube was empty. To reduce droplet shear or breakage of the droplets, a bigger sort nozzle (100 µm) was used. The samples were sorted at a sample pressure with a flow rate of 1-3, a system pressure between 8-10 psi and a drop frequency of ~30 kHz. During the FACS measurements, the samples tubes were regularly agitated by the operator to prevent the DEs from settling. The DEs were gated as described above and the coumarin peak was further split into four gates (P3-P6) according to increasing coumarin fluorescence (P3 = 10%, P4 = ~70%, P5 = 10%, P6 = 5%). The DE sample was sorted into four collection tubes according to these four gates. For a general experiment, the sorting was carried out until a minimum of 20’000 events were reached in the

highest gate (P6). For sorts where the highest gate was set to 1%, all produced DE samples were sorted (generally ~40 min of production yielded ~7000 sorted events in P6 = 1%).

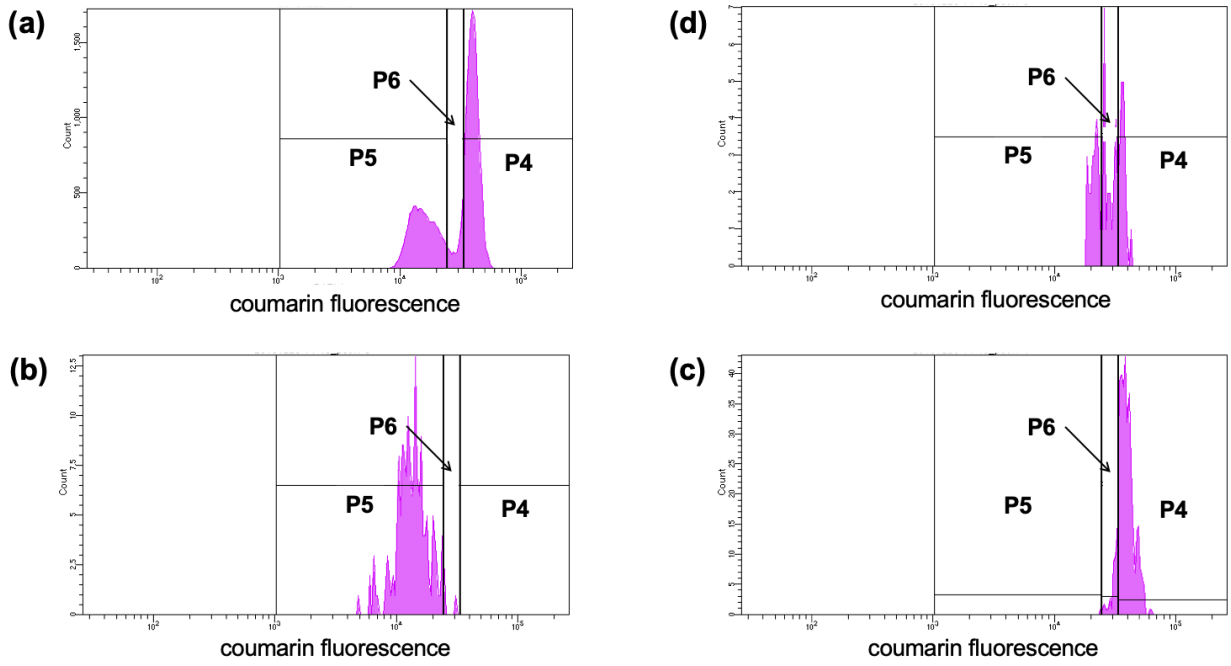


Figure S37. Resorting of different gates. a) Fluorescence intensity distribution of DEs. **b)** Sorted gate P5 remeasured. **c)** Sorted gate P6 remeasured. **d)** Sorted gate P4 remeasured. Efficient sorting of the DEs with minimal overlap was achieved. The resorting also proved that the droplets remained intact during the sorting process.

Table S18. Summary of sort reports indicating the total sorted events.

sorted gate	number of events
control 99:1_P3	75’013
control 99:1_P4	153’092
control 99:1_P5	49’086
control 99:1_P6-5%	20’762
control 99:1_P6-1%	7’649
400-variant library_P3	104’157
400-variant library_P4	325’212
400-variant library_P5	63’617
400-variant library_P6-5%	30’550
160’000-variant library_P3	96’708
160’000-variant library_P4	131’367
160’000-variant library_P5	59’119
160’000-variant library_P6-5%	25’516
160’000-variant library_P6-1%	7’147

C.2.7 DNA SEQUENCING

Table S19. Primers used for PCR amplification for Nanopore Sequencing/NGS.

Primer	Sequence
TOPO_fwd	atatgcgattaccccggaaattgcgaccggagctag
TOPO_rev	gccaaaacagccaagcttgtcgacggagctcgaatt
NGS_adapter_L1	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatcttatcacgagaagcacgcattaataccc
NGS_adapter_L2	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctatcgatgtagaagcacgcattaataccc
NGS_adapter_L3	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgatcttgtaagaagcacgcattaataccc
NGS_adapter_L4	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctcgatgccaatagaagcacgcattaataccc
NGS_adapter_L5	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatcttcgatacagtgagaagcacgcattaataccc
NGS_adapter_R1	caagcagaagacggcatacgagatgtgactggagttcagacgtgtgctcttccgatcttacgttattcggacggcttcaccttg
NGS_adapter_R2	caagcagaagacggcatacgagatgtgactggagttcagacgtgtgctcttccgatctatttcaagaacggacggcttcaccttg
NGS_adapter_R3	caagcagaagacggcatacgagatgtgactggagttcagacgtgtgctcttccgatctgatgatctgccggacggcttcaccttg
NGS_adapter_R4	caagcagaagacggcatacgagatgtgactggagttcagacgtgtgctcttccgatctcgattaacatagcggacggcttcaccttg
NGS_adapter_R5	caagcagaagacggcatacgagatgtgactggagttcagacgtgtgctcttccgatcttcgatgcgtcaaccggacggcttcaccttg
Nanopore_fwd	taatacgactcactatagg
Nanopore_rev	aaaccgtctatcagggc

blue = i5 indices, violet = i7 indices, orange = barcodes, green = primer binding site.

PLASMID RECOVERY

The sorted samples were spun down, and the supernatant (PBS + 0.5% SDS) was removed. The droplets were then resuspended in the resuspension buffer (B1, 50 µL) of the Monarch plasmid extraction kit. Additionally, 1 H, 1 H, 2 H, 2H-perfluorooctanol (12 µL) was added and the suspension was incubated at 50 °C for 20-40 min and 300 rpm. Then, the B2 solution (50 µL) was added and incubated at room temperature for 1 min. Finally, the B3 solution (100 µL) was added and incubated at room temperature for 2 min. The further steps were carried out following the manufacturer's protocol (Monarch PCR clean-up kit by NEB) and elution was performed twice with 6 µL of mQ H_2O (heated to 70 °C).

PCR AMPLIFICATION

Plasmid DNA (extracted from the five sorted populations and the parent) was amplified using either the TOPO cloning primers, NGS primers already containing the barcodes, adapters and indices for sequencing, or the nanopore primers (Table S19). For PCR amplification, 5 µL of the extracted plasmid was mixed with 1 µL of the forward primer, 1 µL of reverse primer and 10 µL of the Q5 Master mix, and the volume was adjusted to 20 µL. The PCR was run with 30-40 cycles at an annealing temperature of 65 °C.

TOPO CLONING AND TRANSFORMATION

The PCR product was used without further purification. For TOPO cloning 4 µL of the PCR mixture was mixed with 1 µL of the salt solution and the volume was adjusted to 6 µL. Following 1 µL of the pCRT^{M}II-Blunt-TOPO® was added and incubated at room temperature for 5-15 min. 2 µL of this solution were transformed into chemically competent cells and plated on LB-agar plates supplemented with kanamycin and incubated overnight at 37 °C. For sequencing, colonies were picked and analyzed by Sanger sequencing at Microsynth AG.

ILLUMINA SEQUENCING

For Illumina sequencing, the PCR products were purified by agarose gel electrophoresis. The concentration and purity were determined by capillary electrophoresis using an Agilent Bioanalyzer 2100. Sequencing was done at the Genomics Facility in Basel, CH and Illumina sequencing using a MiSeq Nano (300 cylces, PE × 150) or a NextSeq Mid Output v2 (300 cycles, PE 2 × 150) kit were done. The different samples were pooled according to the concentrations at ~250 bp and ~800 bp to have an equal molar input for the NGS run. The runs were spiked with additional 20% PhiX. Primary data analyses was done with Illumina RTA version 2.4.11 and bcl2fastq v2.20.0.422.

NANOPORE MEASUREMENTS

The individual PCR products were purified following the manufacturer's protocol (Monarch PCR clean-up kit by NEB) and elution was performed once with 6 µL of mQ H_2O (heated to 70 °C). DNA barcoding was carried out by mixing the 6 µL of eluted plasmid DNA with 1 µL of barcode purchased from Oxford Nanopore Technologies. The barcoding was done according to manufacturer's protocol (Native Barcoding Expansion 1-12 (PCR-free) EXP-NBD104 kit by Oxford Nanopore Technologies). After purification of the individual barcoding reactions using magnetic beads (AMPure XP) the sequencing samples were prepared according to manufacturer's protocol (Ligation Sequencing SQK-LSK109 kit by Oxford Nanopore Technologies) and analyzed using a MinION sequencer in combination with a R9.4.1 pore-based flowcell FLO-MIN106D, also from Oxford Nanopore Technologies.

EXPRESSION AND PURIFICATION OF THE SCDSAV VARIANTS

Mutations were introduced by individually designed primers and PCR was carried out following the QuikChange^{TM} protcol. Expression and purification of scdSav was carried out as described elsewhere. Briefly *E. coli* BL21 (DE3) cells transformed with a pRSF-scdSav plasmid bearing the correspinding mutations were cultivated in ZYP-5052 medium. After harvesting the

cells, the pellet was lysed, and the purification of the proteins was performed using AKTA FPLC with an iminobiotin column (Sepharose™ CL-4B from Affiland, column volume = 50 mL). The pure protein solution was frozen and then lyophilized over two days at -50 °C and 0.12 mbar to obtain the pure protein as a white fluffy solid.

VALIDATION BY IN VITRO SCREENING.

Purified scdSav (pp-scdSav, 2 mM; 10 µM final concentration) was mixed with the cofactor stock solution (1 mM; 5 µM final concentration) in a 96-well plate. The mixture was incubated for 5-10 min before substrate **44** (10 mM; 500 µM final concentration) was added. The plate was incubated at 37 °C and 300 rpm, and measurements were taken at various timepoints (XX h, XX h, XX h, XX h). The coumarin fluorescence was determined in a TECAN plate reader at λ_{ex} = 395 nm and λ_{em} = 460 nm. All measurements were carried out in triplicates (Figure 5f, Figure SXX). The TON was calculated using a standard curve (Figure SXX). The calibration curve was prepared in PBS with different concentrations of substrate and product ranging from 1 µM to 50 µM. (to be completed)

C.2.8 DATA EVALUATION

We used in house bash and R scripts to analyze the NGS data. Fastq files containing the forward and reverse reads were obtained following NGS. The reads were extracted and paired. The reads were filtered using the 24-bp fixed region located between position 112 and position 121, allowing for max. 3 mismatches. The target fragments were attributed to each sample using their unique barcode (Table S19). The mutations at positions 112 and 121 were identified by retrieving the 3 nucleotides just before and just after the fixed sequence. For the scd library, the 3 nucleotides located 23-bp after the fixed region were also retrieved. This allowed the verification of the amino acid at position 128 which was mutated between monomer A and monomer B (alanine for monomer A and arginine for monomer B).

Table S20. Average numbers for NGS and Nanopore sequencing of the 400- and 160'000-variant libraries.

library	event	number of reads
400-variant library *via* NGS	total	113'161
	usable reads	72'323
	Sav-wt	693
160'000-variant library *via* NGS	total	567696
	usable reads	184
	scdSav-wt	15
160'000-variant library *via* NGS	total	3516
	usable reads	128
	scdSav-wt	0

C.2.9 PLASMIDS AND SEQUENCES

PERIPLASMIC SAV

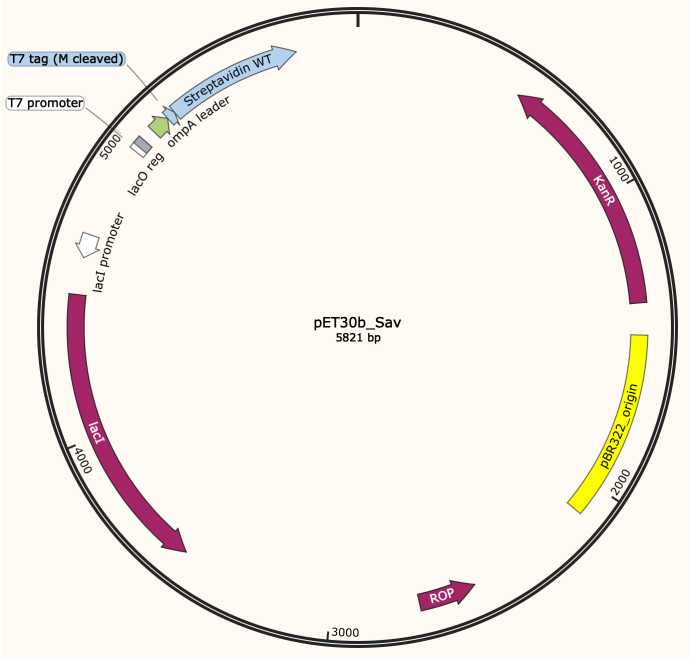


Figure S38. Plasmid map of pET30b encoding scdSav

DNA sequence of periplasmic Sav. red = positions S112 and K121.

ATGAAAAAGACAGCTATCGCGATTGCAGTGGCACTGGCTGGTTTCGCTACCGTA
GCGCAGGCCGCTAGCATGACTGGTGGACAGCAAATGGGTCGGGATCAGGCCGGCATC
ACCGGCACCTGGTACAACCAGCTCGGCTCGACCTTCATCGTGACCGCGGGCGCCGAC
GGCGCCCTGACCGGAACCTACGAGTCGGCCGTCGGCAACGCCGAGAGCCGCTACGTC
CTGACCGGTCGTTACGACAGCGCCCCGGCCACCGACGGCAGCGGCACCGCCCTCGGT
TGGACGGTGGCCTGGAAGAATAACTACCGCAACGCCCACTCCGCGACCACGTGGAGC
GGCCAGTACGTCGGCGGCGCCGAGGCGAGGATCAACACCCAGTGGCTGCTGACCTCC
GGCACCACCGAGGCCAACGCCTGGAAGTCCACGCTGGTCGGCCACGACACCTTCACC
AAGGTGAAGCCGTCCGCCGCCTCCATCGACGCGGCGAAGAAGGCCGGCGTCAACAAC
GGCAACCCGCTCGACGCCGTTCAGCAGTAATAA.

Protein sequence of periplasmic Sav. red = positions S112 and K121.

MKKTAIAIAVALAGFATVAQAASMTGGQQMGRDQAGITGTWYNQLGSTFIVTAGA
DGALTGTYESAVGNAESRYVLTGRYDSAPATDGSGTALGWTVAWKNNYRNAHSATTW
SGQYVGGAEARINTQWLLTSGTTEANAWKSTLVGHDTFTKVKPSAASIDAAKKAGVNN
GNPLDAVQQ.

PERIPLASMIC SCDSAV

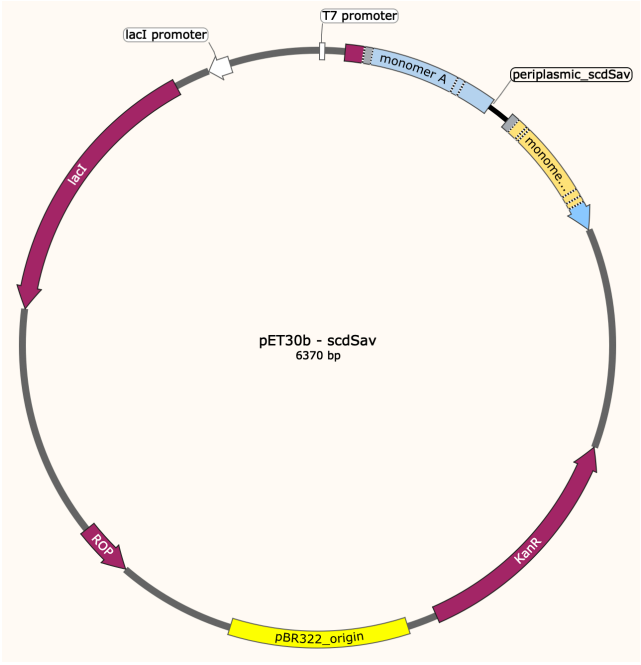


Figure S39. Plasmid map of pET30b encoding Sav.

DNA sequence of periplasmic scdSav. red = positions S112^{A} and K121^{A}, and turquoise = positions S112^{B} and K121^{B}.

ATGAAAAAGACAGCTATCGCGATTGCAGTGGCACTGGCTGGTTTCGCTACCGTA
GCGCAGGCCGCAAGCATGACTGGAGGTCAGCAAATGGGTCGTGATCAGGCAGGTATT
ACCGGCACCTGGTATAATCAACTGGGTAGCACCTTTATTGTTACTGCAGGCGCAGATG
GTGCACTGACCGGTACGTATGAAAGCGCAGTTGGTAATGCAGAAAGCCGTTATGTTC
TGACAGGTCGTTATGATAGCGCACCGGCAACCGATGGTAGCGGCACCGCACTGGGTT
GGACCGTTGCATGGAAAAATAACTATCGTAATGCACATAGCGCAACCACCTGGTCAG
GTCAGTATGTTGGTGGTGCAGAAGCACGCATTAATACCCAGTGGCTGCTGACCAGCG
GCACCACCGAAGCAAATGCCTGGGCAAGCACCCTGGTTGGTCATGATACCTTTACCA
AAGTTAAACCGAGCGCAGCAAGCATTGATGCAGCAAAAAAAGCCGGTGTGAATAAT
GGTAATCCGCTGGATGCAGTTCAGCAGggatccggtggcggtaacggtgggggaaacggtggcggaaatggcg
gagggaacattgatggtcgcggtggtggtaatGCTAGCATGACTGGTGGACAGCAAATGGGTCGGGATC
AGGCCGGCATAACCGGCACCTGGTACGCCCAGCTCGGCGATACCTTCATCGTGACCG
CGGGCGCCGACGGCGCCCTGACCGGAACCTACGTCACGGCGCGTGGCAACGCCGAGA
GCAGATACGTCCTGACCGGTCGTTACGACAGCGCCCCAGCCACCGACGGCTCTGGCA
CCGCCCTCGGTTGGACGGTGGCCTGGAAGAACAATTACAGAAACGCCCACTCCGCGA
CCACGTGGAGCGGCCAATACGTCGGCGGCGCCGAGGCGAGGATCAACACACAATGGT
TATTAACAAAAGGAACTACTGAGGCCAACGCATGGAAGTCCACGCTGGTCGGCTGCG

CCACCTTCACCAAGGTGAAGCCTTCCGCCGCCTCAATCGACGCGGCGAAGAAGGCTG
GCGTCAACAACGGCAACCCTCTCGACGCCGTACAACAATAA.

Protein sequence of periplasmic scdSav. red = positions S112^{A} and K121^{A}, and turquoise = positions S112^{B} and K121^{B}.

MKKTAIAIAVALAGFATVAQAASMTGGQQMGRDQAGITGTWYNQLGSTFIVTAGA
DGALTGTYESAVGNAESRYVLTGRYDSAPATDGSGTALGWTVAWKNNYRNAHSATTW
SGQYVGGAEARINTQWLLTSGTTEANAWKSTLVGHDTFTKVKPSAASIDAAKKAGVNN
GNPLDAVQQGSGGGNGGGNGGGNGGGNIDGRGGGNASMTGGQQMGRDQAGITGTWY
AQLGDTFIVTAGADGALTGTYVTARGNAESRYVLTGRYDSAPATDGSGTALGWTVAWK
NNYRNAHSATTWSGQYVGGAEARINTQWLLTSGTTEANAWKSTLVGCATFTKVKPSAA
SIDAAKKAGVNNGNPLDAVQQ.

CONSTITUTIVELY EXPRESSED MNECTARINE

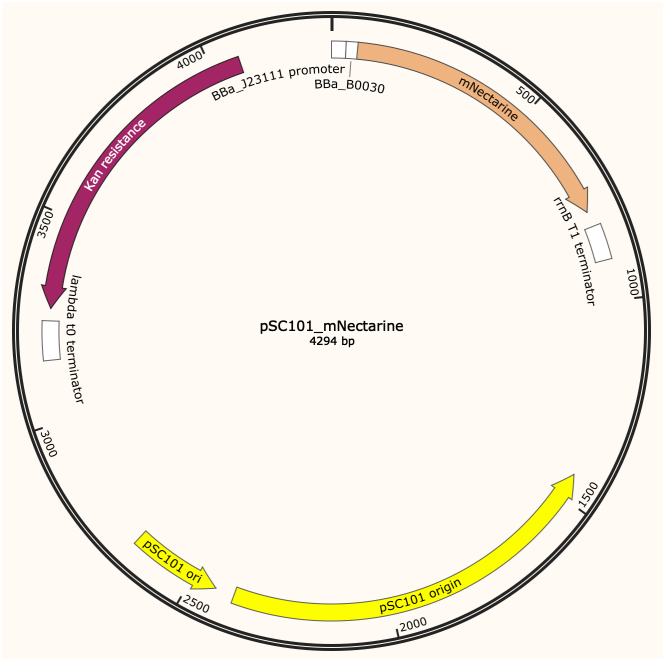


Figure S40. Plasmid map of pSC101 encoding mNectarine.

DNA sequence of mNectarine

ATGGTTAGCAAAGGCGAAGAAGATAACATGGCCATTATCAAGGAGTTTATGCGT
TTCAAAGTTCATATGGAGGGTTCCGTCAACGGGCATGAGTTCGAGATTGAAGGCGAG
GGGGAGGGCCGTCCGTACGAGGGTACACAAACAGCCAAACTGAAAGTCACGAAGGG
TGGACCACTTCCGTTCGCGTGGGATATCCTGTCACCTCAATTCTGCTATGGAAGCAAA
GCGTACGTGAAACATCCGGCCGATATTCCGGACTATCTGAAACTGTCGTTCCCTGAAG

GTTTAAACTGGGAACGTGTGATGAACTTTGAGGACGGCGGGGTTGTTACCGTAACGC
AGGATTCCTCTCTGCAAGACGGCGAGTTCATTTACAAGGTCAAACTCCGTGGTACAAA
TTTTCCCAGTGATGGTCCAGTTATGCAGTGTCGCACCGTTGGTTGGGAGGCCAGTACC
GAACGTATGCATCCGGAGGACGGCGCACTCAAAGGCGAAATCATGCAACGCTTAAAG
CTCAAGGACGGCGGTCATTACGACGCGGAAGTCAAAACAACTTACAAAGCAAAAAA
ACCTGTGCAGTTACCGGGTGCATATAACGTCGACATTAAACTCGATATTCTTTCCCAT
AACGAGGACTACACAATCGTTGAGCTGTACGAACGCGCGGAAGGCCGTCACAGCACG
GGTGGGATGGACGAGCTCTATAAGTAATAA.

Protein sequence of mNectarine

MVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTK
GGPLPFAWDILSPQFCYGSKAYVKHPADIPDYLKLSFPEGLNWERVMNFEDGGVVTVTQ
DSSLQDGEFIYKVKLRGTNFPSDGPVMQCRTVGWEASTERMHPEDGALKGEIMQRLKLK
DGGHYDAEVKTTYKAKKPVQLPGAYNVDIKLDILSHNEDYTIVELYERAEGRHSTGGMD
ELYK.

CONSTITUTIVELY EXPRESSED GFP

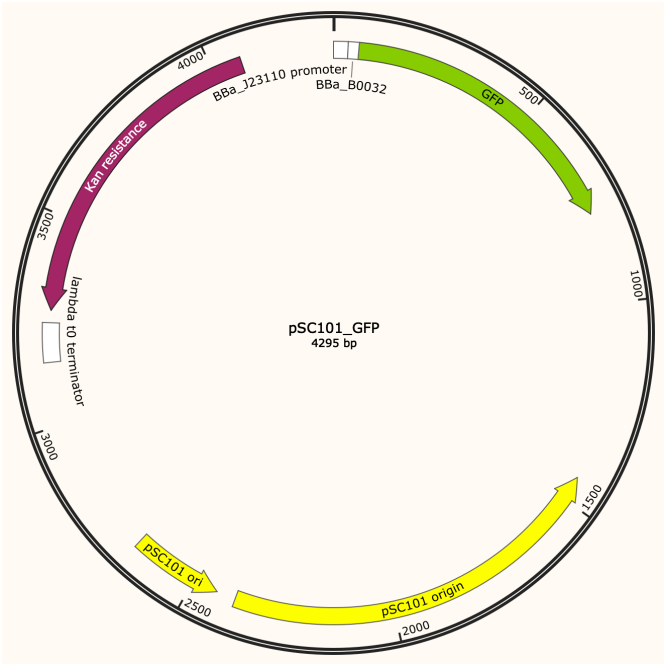


Figure S41. Plasmid map of pSC101 encoding GFP.

DNA sequence of mNectarine

ATGAGTAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTA
GATGGTGATGTTAATGGGCACAAATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCA

ACATACGGAAAACTTACCCTTAAATTTATTTGCACTACTGGAAAACTACCTGTTCCAT
GGCCAACACTTGTCACTACTTTCGCGTATGGTCTTCAATGCTTTGCGAGATACCCAGA
TCATATGAAACAGCATGACTTTTTCAAGAGTGCCATGCCCGAAGGTTATGTACAGGA
AAGAACTATATTTTTCAAAGATGACGGGAACTACAAGACACGTGCTGAAGTCAAGTT
TGAAGGTGATACCCTTGTTAATAGAATCGAGTTAAAAGGTATTGATTTTAAAGAAGAT
GGAAACATTCTTGGACACAAATTGGAATACAACTATAACTCACACAATGTATACATC
ATGGCAGACAAACAAAAGAATGGAATCAAAGTTAACTTCAAAATTAGACACAACATT
GAAGATGGAAGCGTTCAACTAGCAGACCATTATCAACAAAATACTCCAATTGGCGAT
GGCCCTGTCCTTTTACCAGACAACCATTACCTGTCCACACAATCTGCCCTTTCGAAAG
ATCCCAACGAAAAGAGAGACCACATGGTCCTTCTTGAGTTTGTAACAGCTGCTGGGA
TTACACATGGCATGGATGAACTATACAAATAA

Protein sequence of GFP

MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVP
WPTLVTTFAYGLQCFARYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFE
GDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGS
VQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITHGMD
ELYK.

C.3 REFERENCES

[180] J. Serrano-Plana, C. Rumo, J. G. Rebelein, R. L. Peterson, M. Barnet, T. R. Ward, *J. Am. Chem. Soc.* **2020**, *142*, 10617–10623.

[328] J. Zhao, J. G. Rebelein, H. Mallin, C. Trindler, M. M. Pellizzoni, T. R. Ward, J. Am. Chem. Soc. 2018, 140, 13171–13175.

[329] L. Zheng, U. Baumann, J. L. Reymond, Nucleic Acids Res. 2004, 32, DOI 10.1093/nar/gnh110.

[211] T. Völker, F. Dempwolff, P. L. Graumann, E. Meggers, Angew. Chem. Int. Ed. Engl. 2014, 53, 10536–10540.

[301] G. Woronoff, A. El Harrak, E. Mayot, O. Schicke, O. J. Miller, P. Soumillion, A. D. Griffiths, M. Ryckelynck, Anal. Chem. 2011, 83, 2852–2857.

[179] T. Heinisch, F. Schwizer, B. Garabedian, E. Csibra, M. Jeschek, J. Vallapurackal, V. B. Pinheiro, P. Marlière, S. Panke, T. R. Ward, Chem. Sci. 2018, 9, 5383–5388.

[302] S. Deshpande, Y. Caspi, A. E. Meijering, C. Dekker, Nat. Commun. 2016, 7, 10447.

[303] A. Stucki, P. Jusková, N. Nuti, S. Schmitt, P. S. Dittrich, *Small Methods* **2021**, *5*, 2100331.

CURRICULUM VITAE

PERSONAL INFORMATION

Name:	Jaicy Vallapurackal
Date of Birth, Place:	20.03.1993, Vienna, Austria
Address:	Leimenstrasse 44 4051 Basel Switzerland
Nationality:	Austria, (Swiss Residence Permit C)
Telephone:	+41 76 377 12 59
E-Mail:	jaicy.vallapurackal@gmail.com
Orcid ID:	0000-0002-8424-3792

EDUCATION

05.2016 – present expected graduation: 05.2021	**Doctor of Philosophy (PhD) in Chemistry** University of Basel, CH PhD-Thesis: *"Development of a Microfluidics-Based Assay for the Directed Evolution of Artificial Metalloenzymes"* Supervised by Prof. Thomas R. Ward, Department of Chemistry
08.2014 – 03.2016	**Master of Science in Chemistry (MSc)** University of Basel, CH Master-Thesis: *"Genetic Optimisation of an Artificial Metathase Using the Biotin–Streptavidin Technology"* Supervised by Prof. Thomas R. Ward, Department of Chemistry
08.2011 – 07.2014	**Bachelor of Science in Chemistry (BSc)** University of Basel, CH Bachelor project work: *"Expanding the Substrate Scope for the Enantioselective Artificial Suzukiase Using the Streptavidin-Biotin Technology"* Supervised by Prof. Thomas R. Ward, Department of Chemistry Bachelor project work: "*Synthetic Studies Towards the Total Synthesis of Periconianone A*" Supervised by Prof. Karl Gademann, Department of Chemistry
08.2006 – 05.2011	**Matur** Gymnasium am Münsterplatz, Basel, CH Main focus: Latin & Immersion (bilingual Matura in German and English)

PRIZES / AWARDS

08.2020	**Best Oral Presentation Award (Runner-up)** – SCS Fall Meeting, Bern, CH Awarded: CHF 400.- and an article in CHIMIA
06.12.2019	**Beitrag an Nachwuchsveranstaltungen** – PCC–Christmas Symposium
06.2019	**Best Poster Award –** PCC Christmas Symposium, Basel, CH
05.2018	**Best Poster Award –** PCC Christmas Symposium, Basel, CH
06.2015	**Best Poster Award (Second Prize) –** Sustainability Symposium, Nottingham, UK
03. – 07.2015	**Swiss European Mobility Program – Student Mobility for Traineeships** Awarded: CHF 1200.- for an internship in Nottingham, UK

EMPLOYMENT HISTORY

05.2016 – present expected graduation: 05.2021	**University of Basel, Basel, CH** – PhD candidate Supervisor: Prof. Dr. Thomas R. Ward, Department of Chemistry Thesis: "*Development of a Microfluidics-Based Screening Assay for the Directed Evolution of Artificial Metalloenzymes"*
07.2016 – 08.2016	**University of California, Irvine, USA** – Research Exchange Supervisor: Prof. Andy S. Borovik Resulting Publication: *"Artificial Metalloproteins Containing Co_4O_4 Cubane Active Sites"* (Link)

05.2016 – 06.2016	**University of Evry, ISSB Genopole, Paris, FR** – Research Exchange Supervisor: Dr. Philippe Marlière Work included: petri dish assay, construction of auxotrophic *E. coli* strains
09.2015 – 03.2016	**University of Basel, Basel, CH** – Master student Supervisor: Prof. Dr. Thomas R. Ward, Department of Chemistry Thesis: '*Genetic Optimisation of an Artificial Metathase Using the Biotin–Streptavidin Technology'*
03.2015 – 08.2015	**University of Nottingham, UK –** Research Internship Supervisor: Prof. Dr. Anca Pordea Topic: *'Artificial Metalloenzymes Based on Transition Metals'*

PUBLICATIONS

[1] Vallapurackal, J.; Stucki, A.; Liang, A. D.; Dittrich, P. S.; Ward, T. R.
"Single-Round Remodelling of the Active Site of an Artificial Metalloenzyme using an Ultrahigh-Throughput Double Emulsion Screening"
Angew. Chem. Int. Ed. ***2022****, 61, e2022073.*

[2] Baiyoumy A.; Vallapurackal J.; Schwizer F.; Heinisch T.; Kardashliev T.; Held, M.; Panke S.; Ward T. R.
"Directed Evolution of a Surface-Displayed Artificial Allylic Deallylase Relying on a GFP Reporter Protein."
ACS Catal. **2021**, *11* (48), 1070–510712.

[3] Leone, S. D.; Vallapurackal J.; Avsar, S. Y.; Kyropolou, M.; Ward, T. R.; Palivan, C. G.; Meier,W.
"Expanding the Potential of Solvent-Assisted Method to Create Bio-Interfaces From Amphiphilic Block Copolymers."
Biomacromolecules, **2021**, *22* (7), 3005–3016.

[4] Vallapurackal J.
"Towards the Directed Evolution of Artificial Metalloenzymes."
Chim. Int. J. Chem. **2021**, *75* (4), 257–260.

[5] Stucki, A.; Vallapurackal, J.; Ward, T. R.; Dittrich, P. S.
"*Droplet Microfluidics and Directed Evolution of Enzymes: An Intertwined Journey."*
Angew. Chemie Int. Ed. **2021**, doi: 10.1002/anie.202016154. (# shared first author)

[6] Heinisch, T.; Schwizer, F.; Garabedian, B.; Csibra, E.; Jeschek, M.; Vallapurackal, J.; Pinheiro, V. B.; Marlière, P.; Panke, S.; Ward, T. R.
"E. Coli Surface Display of Streptavidin for Directed Evolution of an Allylic Deallylase."
Chem. Sci. **2018**, *9* (24), 5383–5388.

[7] Olshansky, L.; Huerta-Lavorie, R.; Nguyen, A. I.; Vallapurackal, J.; Furst, A.; Tilley, T. D.; Borovik, A. S.
"*Artificial Metalloproteins Containing Co4O4 Cubane Active Sites."*
J. Am. Chem. Soc. **2018**, *140* (8), 2739–2742.

[8] Chatterjee, A.; Mallin, H.; Klehr, J.; Vallapurackal, J.; Finke, A. D.; Vera, L.; Marsh, M.; Ward, T. R.
"An Enantioselective Artificial Suzukiase Based on the Biotin–Streptavidin Technology"

PERSONAL SKILLS

Languages:	German (native), Malayalam (native), English (fluent), French (moderate), Tamil (basic), Hindi (basic), Latin (certificate of proficiency)
Research Skills:	Synthesis of small molecules, UPLC/MS, GC/MS, NMR, *in vitro* & *in vivo* catalysis, protein engineering, western blot, cloning & DNA library creation, directed evolution, ultrahigh-throughput screening, screening in multititer plates, droplet microfluidics, PDMS-chip manipulation, Flow Cytometry, FACS, Nanopore Sequencing, NGS
Digital Competences:	ChemDraw, SciFinder, Reaxys, MestraNova, FlowJo, MS-Office, Snapgene, PyMol, Python, Illustrator, PremierePro, Lightroom, Photoshop
Soft Skills:	Creative, collaborative, team-oriented, written and oral science communication, public outreach, leadership, team management, project management, problem solving, prioritizing & multitasking, conference organization