Diss. ETH No. 23295

Assembly and Properties of Enzyme-Capsid Complexes

A thesis submitted to attain the degree of
DOCTOR OF SCIENCES of ETH ZURICH
(Dr. sc. ETH Zürich)

presented by

REINHARD MATTHIAS ZSCHOCHE
MSc Chemistry, ETH Zurich
born on 19.05.1987
citizen of Germany

accepted on the recommendation of

Prof. Dr. Donald Hilvert
Prof. Dr. Helma Wennemers

2016

The best way to have a good idea is to have a lot of ideas.

Linus Pauling

Acknowledgements

I am extremely grateful to Prof. Donald Hilvert for taking me on as a PhD student and mentoring me throughout this journey. From the start, he granted me the freedom to conceive my own project, encouraged me to think big and pursue a wide variety of ideas. Don's door was always open and with his remarkable attention to detail he every so often nudged my projects into the right direction while motivating by pointing out that the glass was generally half-full. The combination of his supreme standards and a humorous spirit created an outstanding and pleasant environment to learn the scientific trade. A big thank-you goes to Prof. Helma Wennemers for inspiring lectures, fruitful discussions and agreeing to co-examine my thesis. I am grateful to the Stipendienfonds der Schweizerischen Chemischen Industrie for awarding me a scholarship to fund my PhD studies.

Cutting edge research relies on mastering state-of-the art techniques and I was fortunate to collaborate with Prof. Albert Heck and Michiel van der Waterbeemd from the University Utrecht, Netherlands in investigating the assembly of various capsids by native mass spectrometry. At ETH, I want to thank Dr. Marc-Olivier Ebert for helping me to conduct isotope exchange experiments followed by NMR spectroscopy, Peter Tittmann for training me in electron microscopy and the team of the LOC mass spectrometry service lab for going the extra mile in determining protein masses. I am grateful to Prof. Benjamin Schuler and Dr. Andrea Holla at the University of Zurich for allowing me to determine fluorescence quantum yields on their instruments.

The joy of going to work is critically dependent on the co-workers and in this regard the environment in the Hilvert group has spoiled me for life. I was privileged to supervise gifted undergraduate and graduate students, Corinna Landig, Jethro Hemmann and Eva van Rooden and I am grateful for their dedication, creativity as well as hard work. Upon my arrival, Lars Giger introduced me to the aldolase project and throughout Richard Obexer as well as Xavi Garrabou have been terrific colleagues. The capsid subgroup, which grew tremendously during the course of my PhD, has been a supportive community in which tips, tricks and resources were readily shared. For this superb environment, which included enthralling subgroup meetings and adventurous trips to conferences I would like to thank Yusuke Azuma, Tobias Beck, Tom Edwardson, Raphael Frey, Takahiro Hayashi, Shiksha Mantri, Eita Sasaki, Stephan Tetter and Zibi Pianowski.

I want to thank Prof. Peter Kast not only for being a knowledgeable counselor on any matter related to molecular or microbiology, but also for placing me in the best of all labs, F336, and constantly refilling it with marvelous co-workers. Yusuke

Azuma, Adrian Bunzel, Tom Edwardson, Norman Metanis, Cindy Schulenburg, Sabine Studer and An Vandemeulebroucke did not only teach me a variety of specialty techniques, but were a spectacular crowd. Our lab was a compassionate environment when Belgian joie de vivre met de Kölsche Frohnatur, in which middle-eastern hummus was served with Salzwedler tree cake paired with Fondant from Visperterminen, and where Japanese premier-league baseball encountered Scottish climbing talent. The lab could not operate without diligent staff who I want to thank for always being considerate and forthcoming.

Some people are difficult to sort into drawers, and in no particular order I want to thank David for being a great pal in organizing lavish group festivities that made those of the late Roman Empire fade in comparison, senior PhD students Hajo and Clemens for being grandfatherly scientific role-models, not least by inexorably discussing the dichotomy of religion and evolution over lunch at 1 pm, Clément for teaching me flamboyant dance moves despite blatant absence of talent, Shiksha for bringing the colors of the Holi festival to Zurich, Richie for luscious Italian and Japanese food, Yusuke and Raphi for afternoons spent on boards waiting to catch the waves of the Pacific ocean at Ventura beach, Sabine for ample last minute tips for what to cook for dinner, Taka for always being quick with a joke, Adrian for proving that Schlager knows no age, Eita for an unforgettable stag night, Susanne for always knowing if there's snow, Roy for sparking and nourishing my passion for fragrance chemistry, Marcel for subtly laying out the hidden treasures of canton Wallis, Linda and Jurate for being commiserative accomplices in the same cohort, Tom for teaching me idiosyncratic expressions, An for quickly spreading the news, Doug for grounding the lab with the serenity of a cowboy from the dusty plains of the south-west, and Anthony for numerous chatty runs in the Hönggerbergian woods and a brief introduction to British Pub cuisine.

I want to thank my friends and family for all their care throughout my studies and dissertation, in particular both my grandfathers who early on ignited my curiosity for the essence of things. Finally, special thanks go to Mi Thich for her neverending patience and loving support.

Parts of this thesis have been published

Chapter 3

Parts of this chapter have been published in:

Zschoche R, Hilvert D: **Diffusion-Limited Cargo Loading of an Engineered Protein Container**. *J. Am. Chem. Soc.* 2015, **137**:16121–16132.

Chapter 4

Parts of this chapter have been published in:

Azuma Y, Zschoche R, Tinzl M, Hilvert D: **Quantitative Packaging of Active Enzymes into a Protein Cage**. *Angew. Chem., Int. Ed.* 2016, **55**:1531–1534.

Table of contents

Abstract

Co-localization of enzymes that catalyze consecutive reactions in a metabolic pathway is a common theme in nature. Besides coordinated regulation of enzyme activities, such arrangements are believed to confer a metabolic advantage to the host organism by channeling intermediates directly between different active sites and thus preventing their loss to bulk solvent. The principles underlying such systems are currently under investigation. These efforts would benefit from careful kinetic experiments with well characterized assemblies that enable systematic variation of parameters like complex size and individual reaction rates.

Toward that goal, this thesis describes the assembly of different enzymes within the engineered protein shell of the capsid-forming bacterial enzyme, lumazine synthase (AaLS). Chapter 1 provides an overview of the state of the art in enzyme co-localization in vitro as well as in vivo and describes methodology to characterize direct substrate transfer between catalysts. Two major strategies for substrate channeling seem to prevail: utilization of covalent tethering – or at least strong supramolecular interactions – to constrain intermediates in the vicinity of sequentially acting enzyme active sites and, alternatively, generation of densely packed enzyme networks spanning more than 100 nm to increase the likelihood that a metabolite will be processed before it diffuses out of the microdomain.

The computationally designed and laboratory evolved retro-aldolase RA95.5-8, which catalyzes the cleavage of 4-hydroxy-4-(6-methoxynaphthalene-2-yl)butan-2-one (methodol) with a >3 x 10^8-fold rate acceleration (k_{cat}/k_{uncat}), was selected as a model enzyme. The availability of many variants of this enzyme with catalytic parameters spanning a broad range makes it an excellent choice for studying the effect of co-encapsulation on the overall efficiency of cascade reactions. Furthermore, its product, which is fluorescent and thus easily detectable even within complex mixtures, can be exploited as a substrate for diverse downstream enzymes, including alcohol dehydrogenases, aldehyde dehydrogenases and transaldolases. Chapter 2 focused on characterizing the mechanism of RA95.5-8. Like natural class I aldolases, it employs a lysine sidechain as a covalent catalytic functional group. Identification of the rate-limiting step in its multistep mechanism is important for understanding the origins of the catalytic effects achieved, for efforts to improve it by further directed evolution and for its application in cascade reactions. The binding constant of the acetone product was measured as inhibition constant for the retro-aldol reaction. The rate of Schiff base formation between acetone and the reactive lysine as well as its hydrolysis was determined by isotope exchange of $H_2[^{16}O]$ with the [^{18}O]-labeled carbonyl oxygen. Deprotonation and reprotonation at the α-carbon of the imine were assayed by hydrogen isotope exchange observed by nuclear magnetic resonance spectroscopy. All processes were found to proceed

faster than the rate-determining step for methodol cleavage, as does dissociation of the other product 6-methoxy-2-naphthaldehyde. With product release excluded as the slow step in catalysis, carbinolamine formation and carbon-carbon bond cleavage remain as candidates, with the former being more likely given that the intermediate preceding C-C bond fission could not be reductively trapped with a variety of borohydride reagents. Additionally, kinetic profiling of the synthetic aldol reaction by RA95.5-8 revealed substantial inhibition by 6-methoxy-2-naphthaldehyde. The susceptibility of RA95.5-8 toward reversible inactivation by the aldehyde product reflects a flexible active site, tightening of which might improve overall enzyme performance.

The engineered capsid variant AaLS-13 is a promising artificial encapsulation system that exploits electrostatic interactions for cargo loading. Before using AaLS-13 variants to construct aldolase-containing nanoreactors, the principles underlying formation of host-guest complexes between positively charged guests and the negatively charged capsid protein were investigated. As described in Chapter 3, a pair of fluorescent proteins was developed, which allows spectroscopic determination of the extent of encapsulation by Förster resonance energy transfer (FRET). The encapsulation process is generally complete within a second, suggesting low energetic barriers for proteins to cross the capsid shell. Formation of intermediate aggregates upon mixing host and guest in vitro complicates capsid loading at low ionic strength, but can be sidestepped by increasing salt concentrations or diluting the components. Encapsulation of guests is completely reversible, and the position of the equilibrium is easily tuned by varying the ionic strength. These results, which challenge the notion that AaLS-13 is a continuous rigid shell, provide valuable information about cargo loading that will guide ongoing efforts to engineer functional host–guest complexes.

Chapter 3 was concluded with experiments addressing the assembly of AaLS-13 capsids in the cytosol of *E. coli* as well as the subunit composition of AaLS capsid variants in vitro. Given the high local charge density on the capsid subunits, increased salt concentrations, which reduce Coulombic repulsion between fragments, were found to facilitate capsid formation. Purification of heterologously expressed capsid protein at low ionic strength (I = 250 mM, yet higher than physiological ionic strength ~150 mM) yielded capsid fragments exclusively. In contrast, when AaLS-13 subunits were co-expressed with positively charged guest proteins and purified under the same conditions, assembled capsids were found. These findings suggest that positively charged proteins, much like sodium chloride, shield the negative charges of the capsid protein. In collaboration with the group of Albert Heck at the Universiteit Utrecht, native mass spectrometry was employed to deter-

mine the number of subunits in various AaLS variants. Consistent with the dodecahedral structure that had previously been determined by X-ray crystallography, wild-type AaLS was found to contain 60 subunits, whereas AaLS-neg, endowed with four additional charges per monomer, is constructed from 180 monomers, consistent with a T=3 icosahedral capsid. An evolutionary intermediate between AaLS-neg and AaLS-13, AaLS-RR was found as a mixture of 180-mers, 210-mers and 240-mers. Although it has not yet been possible to resolve the charge envelope of AaLS-13, its position in the mass spectrum relative to AaLS-RR indicates that this capsid is composed of more than 240 subunits.

The loading of AaLS-neg and AaLS-13 with enzymes fused to positively charged tags is described in Chapter 4. Supercharged green fluorescent protein was found superior to a deca-arginine peptide for directing a broad variety of enzymes, including the retro-aldolase RA95.5-8, to the lumen of these capsids. However, encapsulation of oligomeric enzymes still poses a challenge. For example, it was not possible to load AaLS-13 capsids with dimeric or tetrameric alcohol dehydrogenases. On the other hand, AaLS-neg capsids proved far more tractable. Nanoreactors were therefore created by co-encapsulating the RA95.5-8 retro-aldolase and a horse liver alcohol dehydrogenase fusion protein in AaLS-neg. Substrate channeling was tested using the competition method, utilizing external aldehyde dehydrogenase to compete with the encapsulated alcohol dehydrogenase for the 6-methoxy-2-naphthaldehyde product formed by the retro-aldolase. Analysis of the metabolite composition in the presence and absence of capsids by ultra-performance liquid chromatography revealed no advantage from co-encapsulation. Although these results argue against an important role for substrate channeling, there is still room for tweaking this system. Employing more proficient enzymes and charged substrates that can interact favorably with the charged capsid shell, for example, may promote substrate channeling.

During our analysis of the AaLS-neg and AaLS-13 encapsulation systems, the question arose how the natural inclusion complex between trimeric riboflavin synthase and 60-meric lumazine synthase, which lacks the engineered electrostatic interactions, assembles in vivo. Experiments addressing this puzzle are described in Chapter 5. Heterologous co-expression of *Aquifex aeolicus* riboflavin synthase and lumazine synthase in *E. coli* was shown to give intact complexes. Inspired by the threefold symmetry shared by the C-termini of the guests and the hydrophobic clefts on the lumenal side of the capsid shell, we tested whether the C-terminal pseudo-coiled-coil of riboflavin synthase might direct cargo proteins to the interior of wild-type AaLS. Indeed, size-exclusion chromatography of superfolder green fluorescent protein tagged with C-terminal peptides derived from riboflavin synthase

revealed association of a single guest protein to wild type AaLS capsids. The resulting complexes are complementary to the engineered cages filled with many cargo molecules and might be potentially useful for stabilizing fragile enzymes.

The strategies developed in the course of this thesis to create and characterize host-guest complexes between enzyme-fusion proteins and protein containers lay the groundwork for constructing functional organelle-like entities. These systems will be invaluable for investigating the effect of enzyme co-confinement on substrate channeling in vitro and eventually might serve as self-contained synthetic biological building blocks for applications in vivo. A hallmark of capsid-guest assemblies based on AaLS-neg and AaLS-13 is the seamless interchange between assembly in vitro and in vivo, so that the exact same complexes can be used for intracellular experiments and detailed characterization outside the cell. Because guest encapsulation has been found to be fully reversible, engineered AaLS capsids are not only promising scaffolds for the development of nanoreactors, but may also serve as effective vehicles for the delivery of therapeutic and diagnostic agents in medicine.

Zusammenfassung

Enzyme, die aufeinanderfolgende Reaktionen eines Stoffwechselweges katalysieren, sind in der Natur häufig in Komplexen organisiert. Dies erlaubt nicht nur die koordinierte Regulierung aller Enzyme, sondern es wird auch vermutet, dass ihre unmittelbare Nachbarschaft dem Organismus einen metabolischen Vorteil verschafft, indem Metabolite direkt zwischen den aktiven Taschen der Enzyme transportiert werden können ohne sie im Zytosol zu verdünnen. Gegenwärtig versuchen Forscher die Voraussetzungen für einen derartigen direkten Metabolittransfer zu verstehen. Diese Problemstellung würde von kinetischen Experimenten mit gut charakterisierten, künstlichen Multienzymkomplexen profitieren, für die man Parameter wie ihre Grösse, Packungsdichte oder die Geschwindigkeit der einzelnen katalysierten Reaktionen systematisch variieren kann.

Zu diesem Zweck beschreibt diese Dissertation den Einschluss verschiedener Enzyme in der umfunktionalisierten Proteinschale des Kapsid bildenden bakteriellen Enzyms Lumazinsynthase (AaLS). Kapitel 1 gibt einen Überblick über den Wissensstand zu Enzymkomplexen – in vitro wie in vivo – und beschreibt Methoden, mit denen der direkte Substrattransfer zwischen Katalysatoren untersucht werden kann. Für die erfolgreiche direkte Substratübertragung scheinen zwei grundsätzliche Strategien zu dominieren. Zum einen können kovalente Bindungen oder starke supramolekulare Wechselwirkungen den Aufenthalt von Zwischenprodukten auf die direkte Umgebung der nacheinander geschalteten Enzyme beschränken. Zum anderen können dicht gepackte Netzwerke von Enzymen, deren Dimensionen häufig 100 nm übersteigen, die Wahrscheinlichkeit erhöhen, dass ein Metabolit enzymatisch verarbeitet wird, bevor er aus einer solchen Enzymmikrodomäne diffundiert.

Retroaldolase RA95.5-8, welche mit Rechner gestützten Methoden entworfen und anschliessend im Labor evolviert wurde, um die Retroaldolreaktion von 4-Hydroxy-4-(6-methoxynaphthyl)-2-butanon (Methodol) mit einer mehr als 3 x 10^8-fachen Beschleunigung der Reaktionsgeschwindigkeit zu katalysieren (k_{cat}/k_{uncat}), wurde als Modelsystem gewählt. Die Verfügbarkeit vieler Varianten dieses Enzyms deren katalytische Parameter ein breites Spektrum abdecken, macht es zu einem geeigneten Kandidaten, um die Auswirkung des Einschlusses in Kapsiden auf die Effizienz von Kaskadenreaktionen zu untersuchen. Ausserdem kann das Produkt der katalysierten Retroaldolreaktion, ein fluoreszierender Napthaldehyd, der leicht in komplexen Mischungen detektiert werden kann, als Substrat für diverse Enzyme verwendet werden, wie zum Beispiel Alkoholdehydrogenasen, Aldehyddehydrogenasen oder Transaldolasen.

Kapitel 2 beschreibt die Charakterisierung des Mechanismus von RA95.5-8. Ebenso wie natürliche Typ-I-Aldolasen verwendet dieses Enzym eine Lysinseitenkette zur kovalenten Katalyse. Die Identifizierung des geschwindigkeitsbestimmenden Schritts dieses aus mehreren Einzelschritten bestehenden Mechanismus ist wichtig um die molekularen Grundlagen der Katalyse zu verstehen, den Katalysator durch gerichtete Evolution weiter zu verbessern, sowie ihn für Kaskadenreaktionen zu verwenden. Die Dissoziationskonstante des Produkts Aceton wurde über dessen Inhibierungskonstante für die Retroaldolreaktion bestimmt. Die Geschwindigkeitskonstante für die Bildung und die Hydrolyse der Schiffschen Base zwischen Aceton und Lysin 83 wurden über den Isotopenaustausch zwischen $H_2[^{16}O]$ und dem $[^{18}O]$-markierten Carbonylsauerstoff ermittelt. Deprotonierung und Reprotonierung am α-Kohlenstoffatom des Imins wurden über Wasserstoffisotopenaustausch mittels Kernmagnetresonanzspektroskopie beobachtet. All diese Schritte und die Dissoziation des zweiten Produkts, 6-Methoxynapthaldehyd, laufen schneller als die gesamte Retroaldolreaktion ab und können deshalb nicht geschwindigkeitslimitierend sein. Nachdem die Dissoziation der Produkte als langsamster Schritt ausgeschlossen werden kann, verbleiben noch die Bildung des Carbinolamins oder die Spaltung der Kohlenstoff-Kohlenstoff-Bindung als mögliche Kandidaten. Da das Intermediat vor der Spaltung der C-C-Bindung nicht reduktiv abgefangen werden konnte, ist es wahrscheinlich, dass dieser Schritt verhältnismässig schnell ist und die Bildung des Carbinolamins somit geschwindigkeitsbestimmend. Die kinetische Analyse der von RA95.5-8 katalysierten synthetischen Aldolreaktion hat ausserdem ergeben, dass die Reaktion stark durch Substratinhibition beeinträchtigt wird. Die Anfälligkeit des Enzyms für die reversible Deaktivierung durch den Aldehyd spiegelt die Flexibilität der aktiven Tasche wider, so dass die Enzymaktivität durch deren Versteifung vermutlich erheblich gesteigert werden könnte.

Das Designerkapsid AaLS-13 ist ein vielversprechender Proteinkontainer, dessen Beladung auf elektrostatischen Wechselwirkungen mit der Fracht beruht. Bevor AaLS-13 zur Herstellung von Aldolase-Nanoreaktoren verwendet werden konnte, musste die Bildung von Gast-Wirt-Komplexen zwischen positiv geladenen Gästen und dem negativ geladenen Kapsid genauer untersucht werden. Wie in Kapitel 3 beschrieben ist, wurde ein Paar fluoreszierender Proteine entwickelt, welches ermöglicht das Ausmass der Kapsidbeladung mittels Förster-Resonanzenergietransfer (FRET) spektroskopisch zu beobachten. Der Einschluss von Gästen in Kapsiden ist generell bereits innerhalb einer Sekunde abgeschlossen, was darauf schliessen lässt, dass Proteine nur geringe Energiebarrieren überwinden müssen, um die Kapsidhülle zu durchqueren. Die zwischenzeitliche Bildung von Proteinaggregaten unmittelbar nach dem Mischen von Wirt und Gast in vitro erschwert die

Beladung der Kapside bei geringer Ionenstärke, kann aber durch Erhöhung der Salzkonzentration oder Verdünnung der Proteinkomponenten vermieden werden. Der Einschluss von Gästen in den Kapsiden ist vollständig reversibel und die Lage des Gleichgewichts kann durch Anpassung der Ionenstärke eingestellt werden. Diese Ergebnisse nähren Zweifel daran, dass AaLS-13 eine kontinuierliche, starre Hülle besitzt, und liefern gleichzeitig wertvolle Einblicke in die Kapsidbeladung, mit deren Hilfe funktionale Gast-Wirt-Komplexe entwickelt werden können.

Der zweite Teil des dritten Kapitels beschreibt Experimente zur Assemblierung des AaLS-13 Kapsids im *E. coli* Zytosol, sowie zur Bestimmung der Anzahl an Untereinheiten verschiedener AaLS Kapsidvarianten. Dabei wurde festgestellt, dass sich angesichts der hohen lokalen Ladungsdichte auf den Kapsiduntereinheiten eine erhöhte Salzkonzentration positiv auf die Kapsidbildung auswirkt, weil sie die Coulomb-Abstossung zwischen den Fragmenten reduziert. Die Aufreinigung von heterolog exprimierten Kapsidproteinen bei niedriger Ionenstärke (I = 250 mM, was die physiologische Ionenstärke von etwa 150 mM immer noch massgeblich übersteigt) ergab ausschliesslich Kapsidfragmente. Wenn AaLS-13 Untereinheiten allerdings gemeinsam mit positiv geladenen Gastproteinen exprimiert wurden, erhielt man nach der Aufreinigung intakte Kapside. Diese Ergebnisse zeigen, dass die negative Ladung der Kapsidfragmente durch positiv geladene Gäste ebenso gut wie durch Natriumchlorid abgeschirmt werden kann. Die Zahl der Untereinheiten pro Kapsid wurde in einer Kollaboration mit der Gruppe von Albert Heck von der Universität Utrecht durch native Massenspektrometrie bestimmt. Im Einklang mit der Kristallstruktur, welche Wildtyp-AaLS als Dodekahedron zeigt, wurden für diese Variante 60 Untereinheiten pro Kapsid festgestellt. AaLS-neg Kapside, welche im Vergleich zu AaLS-wt vier zusätzliche negative Ladungen pro Monomer aufweisen, bestehen hingegen aus 180 Untereinheiten, was einem ikosahedral-symmetrischen Kapsid mit Triangulationszahl T=3 entspricht. Für AaLS-RR, eine evolutionäre Zwischenstufe zwischen AaLS-neg und AaLS-13, wurde eine Mischung aus 180-meren, 210-meren und 240-meren gefunden. Obwohl es noch nicht gelungen ist, das Massenspektrum von AaLS-13 eindeutig einer Partikelgrösse zuzuordnen, deutet die Position der Signale auf ein Kapsid hin, das aus zwischen 240 und 420 Untereinheiten besteht.

Die Beladung von AaLS-neg und AaLS-13 mit Enzymen, die mit positiv geladenen Peptiden oder Proteinen versehen wurden, ist in Kapitel 4 beschrieben. Um eine Bandbreite von Enzymen, inklusive der Retroaldolase RA95.5-8, ins Innere der Kapside zu dirigieren, wurde eine superpositiv geladene Variante des grünfluoreszierenden Proteins für geeigneter befunden als ein Deka-Argininpeptid. Nichtsdestotrotz stellt die Beladung mit oligomeren Enzymen immer noch eine Herausforderung dar. Beispielsweise war es nicht möglich AaLS-13 Kapside mit dimeren oder

tetrameren Alkoholdehydrogenasen zu beladen. AaLS-neg war vergleichsweise einfacher zu handhaben, weshalb Nanoreaktoren durch den gemeinsamen Einschluss von Fusionsproteinen mit RA95.5-8 und einer Alkoholdehydrogenase in diesem Kapsid hergestellt wurden. Die direkte Übertragung des Metaboliten zwischen den beiden Enzymen wurde mit Hilfe der Kompetitionsmethode untersucht, wobei externe Aldehyddehydrogenasen mit den eingeschlossenen Alkoholdehydrogenase um das von der Retroaldolase gebildete 6-Methoxynapthaldehyd-Zwischenprodukt kompetierten. Die Analyse der Reaktionsmischungen mit Flüssigkeitschromatographie hat gezeigt, dass der gemeinsame Einschluss der Enzyme in ein Kapsid für diesen Versuchsaufbau keinen Vorteil bringt. Obwohl dieses Ergebnis gegen eine entscheidende Rolle für den direkten Metabolittransfer spricht, ist das Potential für die Optimierung dieses Systems noch nicht ausgeschöpft. Der Einsatz schnellerer Enzyme und geladener Substrate, welche vorteilhaft mit der geladenen Kapsidhülle interagieren sollten, könnte die direkte Substratübertragung begünstigen.

Während der Untersuchung der künstlichen AaLS-neg und AaLS-13 Beladungssysteme, kam die Frage auf, wie der natürliche Komplex zwischen trimerer Riboflavinsynthase und 60-mere Lumazinsynthase ohne gezielt eingebaute elektrostatische Wechselwirkungen in vivo gebildet wird. Experimente zur Lösung dieses Rätsels sind in Kapitel 5 beschrieben. Die heterologe gemeinsame Expression von Riboflavinsynthase und Lumazinsynthase aus dem hyperthermophilen Bakterium *Aquifex aeolicus* in *E. coli* hat zur Bildung intakter Komplexe geführt. Angeregt durch die dreifache Rotationssymmetrie, die sowohl am C-Terminus der Gastproteine, wie auch in der Anordnung der hydrophoben Bindetaschen auf der Innenseite der Kapsidhülle gefunden wird, legten nahe, dass der C-terminale Pseudo-Coiled-Coil der Riboflavinsynthase diese in das Innere des Lumazinsynthasekapsids dirigiert. In der Tat konnte mittels Grössenausschlusschromatographie gezeigt werden, dass grünfluoreszierendes Protein, welches mit dem C-terminalen Peptid der Riboflavinsynthase versehen wurde, in AaLS-wt eingeschlossen wird. Die Eigenschaften dieser Komplexe sind von denjenigen, die mit den modifizierten, negativ geladenen Kapsiden gebildet werden, grundsätzlich verschieden. Der Einschluss in stabile AaLS-wt Kapside könnte für die Stabilisierung empfindlicher Enzyme hilfreich sein.

Die Strategien zur Bildung und Charakterisierung von Gast-Wirt-Komplexen zwischen Enzymfusionsproteinen und verschiedenen Kapsiden, die im Zuge dieser Dissertation entwickelt worden sind, legen den Grundstein für die Entwicklung funktionaler künstlicher Zellorganellen. Solche Systeme werden wertvolle Beiträge für die Untersuchung von Multienzymkomplexen und der Auswirkung des gemeinsamen Einschlusses von Enzymen auf den Metabolittransfer leisten. Schlussendlich könnten solche Komplexe als eigenständige Komponenten für Anwendungen in der

synthetischen Biologie verwendet werden. Ein Alleinstellungsmerkmal von AaLS-neg und AaLS-13 ist die Bildung identischer Gast-Wirt-Komplexe in vitro und in vivo, so dass die gleichen Systeme ausgiebig ausserhalb der Zelle charakterisiert und für intrazelluläre Experimente verwendet werden können. Weil ihre Beladung reversibel ist, sind diese Kapside nicht nur geeignete Kandidaten für die Entwicklung von Nanoreaktoren, sondern können auch als molekulare Transporter von therapeutischen und diagnostischen Substanzen in der Medizin Anwendung finden.

1 Introduction

1.1 Toward sustainability with biotechnology

The technological progress achieved in the 20th century was the greatest ever witnessed in the history of humankind. Within less than 100 years, we advanced from a civilization that relied on animals to till fields to a society characterized by petroleum-fueled individual transport, personal computational devices and heavily automatized manufacturing. This growth has been accompanied by an unprecedented use of natural resources and an unparalleled production of waste and environmental pollutants. In their seminal 1972 book on the Limits of Growth, the members of the Club of Rome warned that due to the finiteness of space as well as natural resources it will be impossible for the exponential economic growth seen in the preceding decades to continue indefinitely [1]. Regularly occurring financial crises aside, the world economy has continued to grow ever since. Nevertheless, rising raw material prices and the increasing visibility of climatic change have stimulated awareness for limitations on resources and put the word *sustainability* on every political and economic agenda. Whereas the 20th century economy was driven by insatiable growth, the dominating driver of the 21st century will be continuous improvements in efficiency. Maintaining current standards of living in the industrialized world, while at the same time, improving those in developing countries will only be possible if we manage to produce more while consuming less.

Undoubtedly, biotechnology will be pivotal in realizing these efficiency improvements, a role that is acknowledged by proclaiming it as the centerpiece of the sixth and current Kondratieff cycle [2]. Biotechnology in the broadest sense is the intentional use of biological agents to human benefit. Its applications range from agriculture (green biotechnology), aquatic systems (blue) and medicine (red) to the chemical and food industry (white). While organisms are employed for technical purposes since millennia, the current biotechnology boom bases upon our relatively recent ability to manipulate DNA, which tremendously facilitates the engineering of organisms toward specific tasks. In line with the global trend towards sustainability, the chemical industry, primarily concerned with the interconversion of molecules, is facing increasing pressures from rising raw material and energy prices as well as environmental regulations. Owing to the development of extensive methodology, almost any organic molecule can be produced synthetically. However, the need for sophisticated transition metal catalysts and reagents with poor atom economy make many processes too inefficient to be industrially viable.

Drawing from the living world to improve efficiency is self-evident. Over billions of years, organisms have participated in an unending race for resource efficiency with the species currently populating the earth being the temporary winners of this

competition. Accumulation of myriads of tiny improvements has generated organisms that are adapted to maximize their fitness in a broad variety of environments. However, organisms have not evolved to supply the energy to drive the industrialized world inhabited by more than 7 billion humans. In this regard, hopes that biotechnology will be a panacea for the entirety of global challenges have to be relativized, as it currently seems unlikely that energy produced from biomass will be able to satisfy the world's energy demand [3]. Unsuited to supply energy, the greatest potential of white biotechnology lies in its implementation for the production of chemicals, substituting resource-intensive processes with more efficient ones and giving access to new sustainable materials that can replace suboptimal current substances.

1.2 Biological constituents in the chemical industry

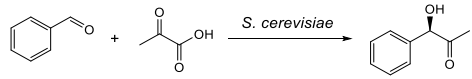
The elegance of biotechnological solutions lies in their ability to self-replicate, using exclusively renewable resources, such as plant biomass, as their feedstock. In contrast, the IT industry relies on a limited stock of precious rare-earth metals while the petrochemical industry requires a continuous supply of fossil oil.

The facile multiplication of biotechnological solutions comes at the expense of high initial development costs, which are a result of the inherent complexity of biological systems. Whereas gas turbines or electronic circuit boards are entirely constructed from principles discovered or invented by humans and are thus fully understood by a person skilled in the art, despite significant progress and pivotal scientific breakthroughs in the recent decades, biological systems are still too complex to be completely described. Even a single isolated entity, a protein, is too complex for us to predict its 3D-dimensional structure ab initio. Systems that are comprehensively understood can be easily manipulated, but one has to resort to trial-and-error when it comes to biological systems whose complexity still exceeds the aptitude of our mind.

1.2.1 Natural systems

Without any knowledge about their mode of action, microorganisms have been used for millennia for the production of alcoholic beverages as well as fermented dairy and cereal products [4]. Fermentation describes the conversion of a feedstock, for example glucose, into a desired product by living microorganisms, with the feedstock usually being initially processed through the primary metabolism (Figure 1). Examples include the production of primary metabolites, such as ethanol, acetic acid, citric acid, amino acids, carbohydrates; secondary metabolites, such as antibiotics; and also peptides as well as proteins. Biotransformation, in contrast, refers to

the conversion of particular substrates into products by a limited number of enzymes. For that matter, the relevant enzymes can be part of an entire cell (whole cell biotransformation) or isolated entities. Conversion of ethanol to acetate aside, one of the first industrially used biotransformations was the decarboxylative benzoin condensation between benzaldehyde and pyruvate using *Saccharomyes cerevisiae* (Scheme 1) [5]. The first example of a crudely purified isolated enzyme is probably invertase, catalyzing the hydrolysis of sucrose to α-D-glucose and β-D-fructose. As a replacement for the sulfuric acid catalyzed process, the filtered supernatant of yeast cells was immobilized on bone char during World War II [4].



Scheme 1. The decarboxylative benzoin condensation between benzaldehyde and pyruvate catalyzed by *Saccharomyces cerevisiae* was one of the first industrially used biotransformations [4,5].

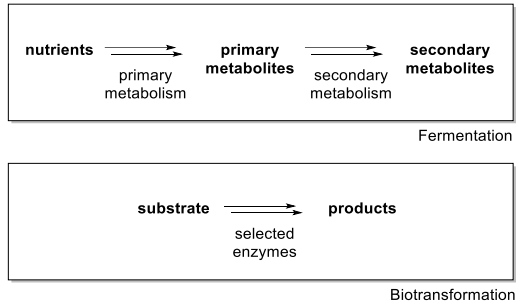


Figure 1. Difference between fermentation and biotransformation.

1.2.2 Engineered systems

During the second half of the 20th century, manifold processes have been developed that exploit microorganisms and their products for industrial chemical processes. Nevertheless, natural systems were often unsuited for application, and with the advent of recombinant DNA technology in the 1980s, people started to manipulate microorganisms and, soon after, the structure of individual proteins. One of the first examples of an industrially relevant enzyme produced by an engineered microorganism was the heterologous expression of thermostable maltogenic α-amylase from *Bacillus stearothermophilus* in *Bacillus subtitilis,* described by researchers of Novo Industri in 1984 [6]. A year later, David Estell, James Wells and co-workers reported mutated variants of the protease subtilisin. Changing a methionine

near the active site to serine or alanine by cassette mutagenesis increased the enzyme's resistance to oxidation by hydrogen peroxide [7].

These two examples illustrate the two main strategies used for engineering biological agents: i) a bottom-up approach in which a single functional unit, usually an enzyme, is subtly modified, often guided by high-resolution three-dimensional structures; and ii) a top-down strategy in which a small number of genes is either introduced into or removed from an entire organism (Figure 2). Both strategies have been continually optimized over the last two decades. Adaptation of enzymes to industrially optimal process conditions with regard to temperature, pH, co-solvent, shear forces, salinity, etc. by directed evolution has become a standard tool [8]. At the same time, entire foreign metabolic pathways have been introduced into microorganisms for the production of complicated secondary metabolites, such as terpenoids [9] and alkaloids [10] in yeast.

As illustrated in Figure 2, bottom-up as well as top-down procedures are well established today. As a consequence, academic interest has begun to focus on the design of completely new enzymes [11,12] and higher-order structures [13,14] from scratch. With increasing experimental evidence corroborating the idea that enzymes in major metabolic pathways are organized into larger supramolecular assemblies, called metabolons [15], synthetic biologists have become receptive to the idea that metabolic flux might be limited more often than not by mass transport. As a consequence, biotechnologists have begun to design artificial scaffolds to organize the biocatalysts of engineered metabolic pathways into multi-enzyme domains (Figure 2, central orange quadrant) with the intention of accelerating overall flux through biocatalytic networks both in vivo and in vitro. Ultimately, such efforts could lead to the creation of entirely new organelles with interiors separated from the cytosol. Such compartmentalization could endow industrially fermentable microorganisms with novel functionalities. On one hand, artificial organelles might facilitate transfer of pathways from higher organisms to single-cellular industrial hosts. On the other hand, they may enable implementation of entirely new chemistries in living cells that have no precedent in nature and might not be attainable if not spatially separated from central cellular metabolism.

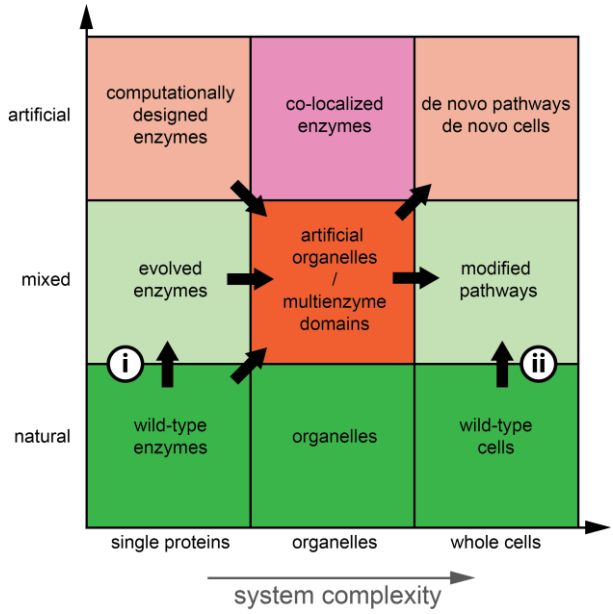


Figure 2. Relationship between natural and artificial agents in white biotechnology. Modified enzymes and organisms (pale green) are easily obtained nowadays by introduction of subtle changes into wild-type species (green; i) and ii)). Co-localization of enzymes in vitro is equally established (pink). Major breakthroughs in the generation of artificial enzymes from scratch have been achieved quite recently through a combination of computational design and directed evolution (pale orange, top left corner). Equally recent are attempts to improve metabolic flux and skew metabolite flow through critical branch points by co-localizing enzymes (orange, middle). The establishment of such microdomains could facilitate installation of completely artificial metabolic pathways or aid the design of entirely artificial cells (pale orange, top-right corner).

Conceptually, artificial organelles can be designed either by de novo bottom-up approaches or by deconstructing and repurposing existing natural entities (Figure 3). With regard to the former, a variety of natural and non-natural scaffolds have been investigated that reversibly or irreversibly co-localize catalysts in vivo and in vitro. Complementarily, cellular organelles and protein-bounded bacterial microcompartments have been modified to incorporate foreign biocatalysts [16,17].

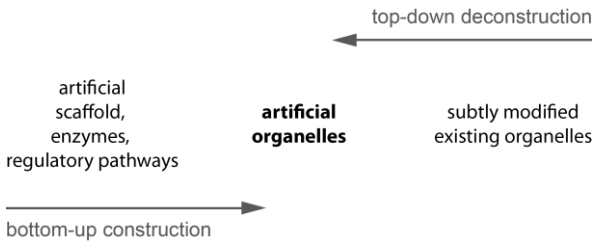


Figure 3. Artificial organelles can in principle be constructed in a bottom-up or top-down fashion.

Although the potential advantages of co-localizing sequentially acting enzymes seem intuitive, they are potentially deceptive. The conditions that need to be fulfilled for directly transferring metabolites between enzyme active sites without having them escape into bulk solution are rather stringent and often not met in designs that proclaim to evince channeling. To acquaint the reader with general approaches for investigating cascade reactions, the methodology used to detect and analyze substrate channeling will be briefly reviewed. Subsequently, findings on natural multi-enzyme complexes will be summarized, followed by an overview of recent efforts to design artificial multi-enzyme structures in vitro as well as in vivo.

1.3 Multi-enzyme cascade reactions

1.3.1 The rate of unhindered diffusion

When describing the purported benefits of cascade reactions, it is often imagined that simply bringing two catalysts closely together will provide the second catalyst in the sequence an advantage due to the high local concentrations of its substrate generated by the first enzyme in its vicinity. Using the example of co-immobilized glucose oxidase and horseradish peroxidase on a DNA origami tile [18], the magnitude of this effect can be estimated using the Einstein-Smoluchowski theory of random Brownian motion [19]. Solving the diffusion equation (Fick's second law, eq. (1)) for the spherically symmetric case yields equation (2), where n(r,t) is the time-dependent spatial number density of Brownian particles (particles per volume), N the number of total particles in the system, D the diffusion coefficient (with unit $\mu m^2\ s^{-1}$), t time, and r the distance from the point of origin for all particles at t = 0.

$$\frac{\partial n(r,t)}{dt} = D\Delta n(r,t) \tag{1}$$

$$n(r,t) = \frac{N}{(4\pi Dt)^{3/2}} exp\left\{-\frac{r^2}{4Dt}\right\} \quad (2)$$

Since all particles are located in an infinitesimally small volume at r = 0 and t = 0, the initial particle density is infinitely high. This situation is described mathematically by a Dirac delta function. The resulting probability distribution for a single particle, such as H_2O_2, with a diffusion coefficient of D = 1,000 µm^2 s^{-1} is shown in Figure 4 as a function of time. Within 10 ms, the time between two turnovers of a fast enzyme (k_{cat} = 100 s^{-1}), the likelihood of finding the product molecule is approximately equal anywhere within a sphere with a 2 µm radius around the enzyme active site. As a reference, the average distance between two molecules at a concentration of 1 nM is about 1.2 µm.

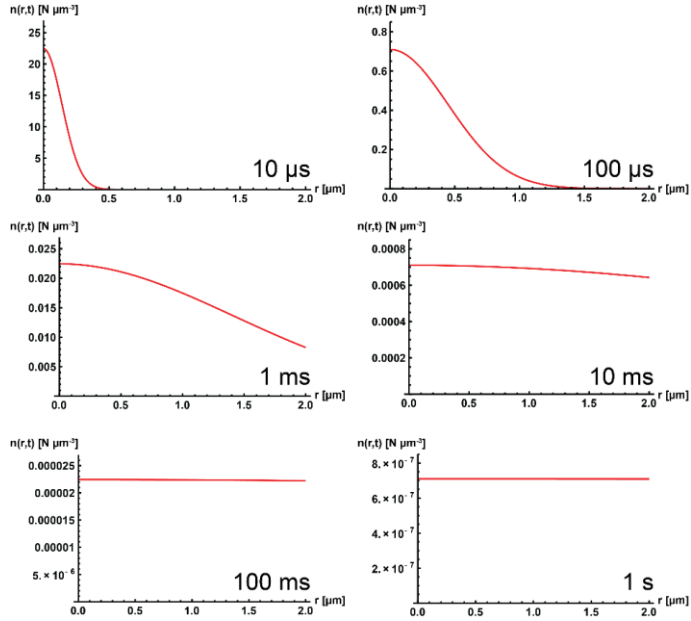


Figure 4. Statistical distribution of a particle with a diffusion coefficient of 1,000 µm^2 s^{-1} after different integration times for a spherically symmetric case.

Yan et al. used summation over each molecule generated in the interval [0,t] with equation (3) to describe the spatial distribution of enzyme products that are continuously generated [18]. τ is the average time between turnovers, i.e. τ = k_{cat}^{-1}.

The number of molecules that remain within a sphere of radius R around the active site within a time t is calculated by eq. (4), and the local average concentration within this sphere is calculated by eq. (5). For k_{cat} = 300 s^{-1} (at saturating substrate concentrations) and R = 10 nm, the resulting distributions are plotted as a function of time (Figure 5).

$$n(r,t) = \sum_{i=0}^{i=\frac{t}{\tau}-1} \frac{N}{(4\pi D(t-i\tau))^{3/2}} exp\left\{-\frac{r^2}{4D(t-i\tau)}\right\} \quad (3)$$

$$N(R,t) = \int_0^R n(r,t) \cdot 4\pi \cdot r^2 dr \quad (4)$$

$$c(R,t) = N(R,t) \cdot \frac{1}{\frac{4}{3}\pi R^3} \cdot \frac{1}{N_A} \quad (5)$$

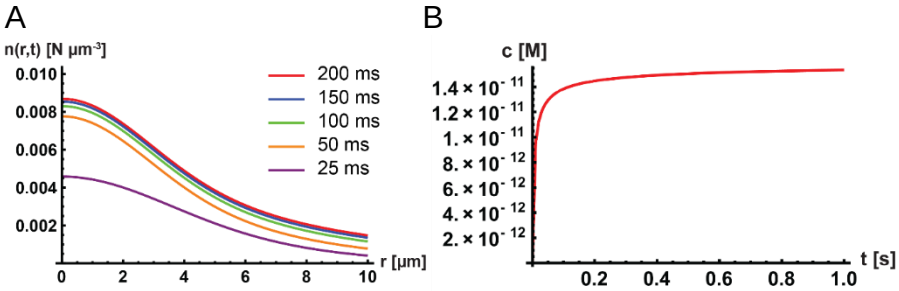


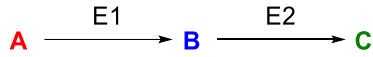
Figure 5. Speed of diffusion. A) Radial distribution of product around an enzyme active site (k_{cat} = 300 s^{-1}, D = 1,000 µm^2 s^{-1}) for different times after reaction initiation. B) Average concentration of product originating from one enzyme active site in a sphere (radius 10 nm) around it as a function of time after reaction initiation.

As is apparent from the graph, after 100 ms a steady-state situation is reached in which the rate of product generation approximately equals the rate of escape by diffusion. The average concentration of product at steady state is less than 15 pM within a 10 nm-radius sphere around the active site. For 1 nM enzyme (k_{cat} = 300 s^{-1}), the concentration of product in bulk solution is already 30 nM after 100 ms. The concentration of product within 10 nm of the active site is only 0.05 % higher, and this difference decreases with longer reaction times. Consequently, high local concentrations of a key intermediate cannot possibly account for substrate channeling in proximal active sites under a free-diffusion regime. This calculation rationalizes why simply fusing two sequentially acting enzymes with a flexible linker does

not improve overall flux without an additional mechanism for substrate channeling in place. Arguably, the diffusion coefficient of 1,000 $\mu m^2 s^{-1}$ for H_2O_2 represents the upper limit for metabolites. However, even 200-fold slower, which might be achieved by increased viscosity, crowding or using larger metabolites, would increase the local concentration only by about 40 nM. As a point of reference, the diffusion coefficient of GFP in *E. coli* cytoplasm is estimated to be 7 $\mu m^2 s^{-1}$ [20].

1.3.2 Kinetic analysis of substrate channeling

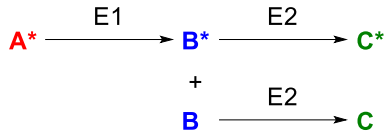
Substrate channeling refers to the coupling of two or more enzymatic reactions in which the common intermediate does not escape into bulk solution, but is directly transferred between the enzyme active sites [21]. Scheme 2 depicts a generic two-enzyme system that will be used to discuss this phenomenon in general terms. Experimental methods to investigate substrate channeling have been comprehensively reviewed by Spivey and Ovádi [21], who focused on natural systems, and more recently by Wheeldon et al. [22], who emphasized man-made scaffolds.



Scheme 2. A generic cascade reaction.

1.3.2.1 Isotope dilution or enrichment

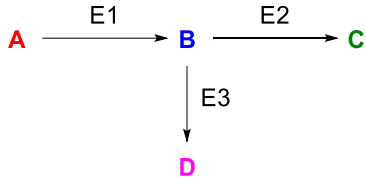
The classic method for detecting substrate channeling is isotope dilution [23]. An isotopically labeled substrate A* is transformed to a product C* via an intermediate B*. The reaction mixture is spiked with non-isotopically labeled B and the ratio of product C with and without isotope-label is determined (Scheme 3). Provided there are no kinetic isotope effects, the product C will be completely isotope-labeled in case of 100 % channeling. In the absence of channeling, the label on C will be diluted by the ratio of B and the steady-state-concentration of B*. In the case of partial channeling, the degree of isotope labeling will be proportional to the degree of channeling. John and Fasella were among the first to use this method in quench flow experiments to demonstrate that oxaloacetate produced by aspartate aminotransferase does not equilibrate with bulk solvent before being reduced by malate dehydrogenase [24].



Scheme 3. Isotope dilution experiment. The asterisk signifies an isotopic label. Whether substrate A or the exogenously added intermediate B is isotope labeled is irrelevant and can be chosen for convenience.

1.3.2.2 Competing Reactions

In another approach to test substrate channeling, the intermediate formed and processed by the E1/E2 complex can be captured by a third enzyme, E3 (Scheme 4). Enzyme E3 must accept B as a substrate, convert it to a different product than E2, and should not interact with either E1 or E2. Many variations of this experiment have been described (examples can be found in references [21] and [22]). A particularly straightforward setup involves testing whether the C/D ratio increases when the reaction is started by addition of A rather than B.



Scheme 4. Channeling of the intermediate B can be detected by addition of a competing enzyme E3 to the reaction mixture.

1.3.2.3 Transient times

Simple cascade reactions without competing reactions and involving only compounds A, B, C and enzymes E1, E2 can be described by differential equations (6)-(8), assuming ideal Michaelis-Menten behavior. Integration of these equations yields the time course of the concentrations of A, B and C (Figure 6).

$$\frac{d[A]}{dt} = -\frac{k_{cat}^{(1)}[E1][A]}{[A] + K_m^{(1)}} \tag{6}$$

$$\frac{d[B]}{dt} = \frac{k_{cat}^{(1)}[E1][A]}{[A] + K_m^{(1)}} - \frac{k_{cat}^{(2)}[E2][B]}{[B] + K_m^{(2)}} \tag{7}$$

$$\frac{d[C]}{dt} = \frac{k_{cat}^{(2)}[E2][B]}{[B] + K_m^{(2)}} \tag{8}$$

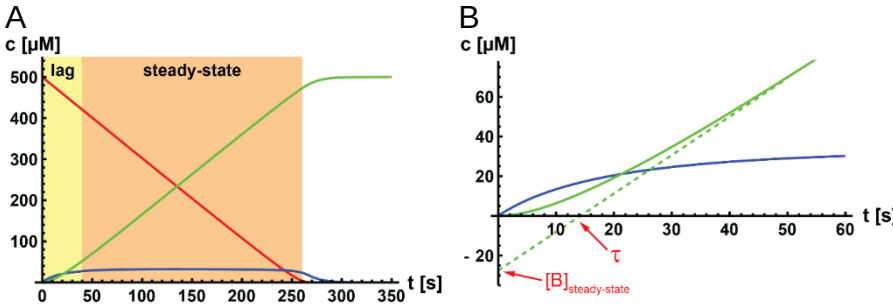


Figure 6. Time-dependent changes in the concentration of metabolites A (red), B (blue) and C (green) (Scheme 2) calculated by integrating equations (6)-(8) using $k_{cat}^{(1)}$ = 200 s^{-1}, $K_m^{(1)}$ = 5 µM, $k_{cat}^{(2)}$ = 500 s^{-1}, $K_m^{(2)}$ = 50 µM, [E1] = [E2] = 10 nM, $[A]_0$ = 500 µM. A) Overall time course. B) Enlarged initial phase. The intercepts of the linear fit (dashed line) for the time dependence of [C] in the steady state phase with x- and y-axes indicate the transient time τ and steady-state concentration of B, respectively. Figures were prepared by plotting the numerical solutions of differential equations (6)-(8) using Mathematica [25]. Figure B was adapted from reference [26].

The phase before the steady-state, the transient phase, is characterized by an increase in [B] until a concentration is reached at which the rate of its consumption by E2, which is a function of [B], equals the rate of its production. If there is channeling of B between the active sites of E1 and E2, this steady-state concentration would be reached earlier at the active site of E2 than in the bulk solution, which would result in decreased transient time. The transient time can be estimated by eq. (9) [21], in which $v^{(1)} = k_{cat}^{(1)} * [E1]$ designates the zero-order rate at which B is formed when E1 is saturated by A. This equation can be extended to include channeling. Under the assumption that [B] is low enough to fulfill subsaturating conditions for E2, it can be written as eq. (10) [27], where p_c is the channeling probability, i.e. the fraction of B that is transferred from E1 to E2 without escaping into bulk solvent, and p_r is the probability that the Michaelis complex [E2•B] proceeds towards C, i.e. $p_r = k_{cat}/(k_{cat} + k_{off})$. The product $p_c \bullet p_r$ for the dihydrofolate reductase-thymidylate synthase two-enzyme complex (see section 1.4, page 36) was estimated to be ≥ 0.78 by comparing the experimentally determined transient time for the catalyzed cascade reaction with the transient time calculated assuming no channeling [27]. If the concentration of B can be monitored during the reaction, channeling can also be detected by a decrease in its steady-state concentration. The

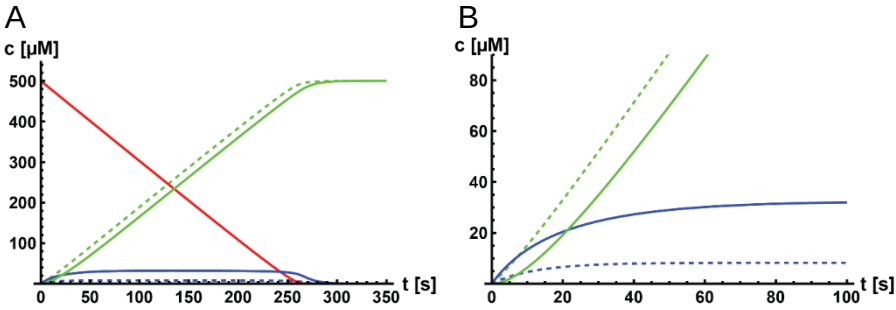


Figure 7. Overview (A) and enlarged initial phase (B) of the time course of metabolite concentration for A (red), B (blue) and C (green) (Scheme 2) calculated by integrating equations (6)-(11) with (dashed lines, $p_c = p_r = 0.8$) and without (solid lines) channeling under the same conditions as for Figure 6. As can be inferred from the identical slopes, intermediate channeling does not increase the overall rate of product formation in the steady state (green lines).

steady-state concentration of intermediate B as a function of channeling and reaction probability can be calculated by equation (11) (for a derivation see reference [27]). For $p_c = p_r = 0.8$ and the conditions described in the legend of Figure 6 the steady-state concentration of B would drop from 33.3 µM to 8.4 µM if the intermediate were channeled between active sites (Figure 7). The steady state concentration of the channeled intermediate calculated by equation (12) is 2.56 nM, which represents only a minute fraction of the total steady-state concentration of the intermediate in eq. (13).

$$\tau = \frac{K_m^{(2)}}{V_{max}^{(2)} - v^{(1)}} \tag{9}$$

$$\tau = \frac{K_m^{(2)}(1 - p_c p_r)}{V_{max}^{(2)}} \tag{10}$$

$$[B]_{steady-state}^{free} = \frac{K_m^{(2)} \cdot v^{(1)}(1 - p_c p_r)}{V_{max}^{(2)} - v^{(1)}(1 - p_c p_r)} \tag{11}$$

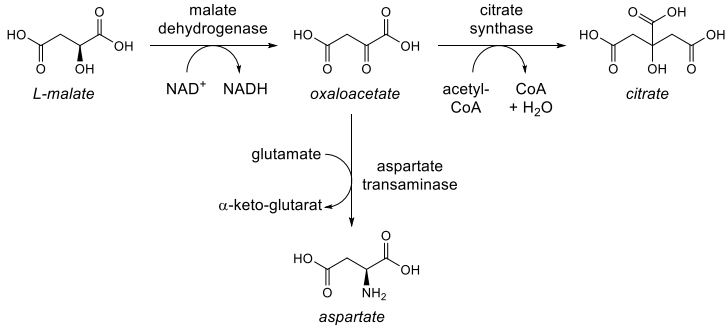
$$[B]_{steady-state}^{channeled} = \frac{v^{(1)} p_c}{k_{cat}^{(2)} + k_{off}^{(2)}} \tag{12}$$

$$[B]_{steady-state}^{tot} = [B]_{steady-state}^{free} + [B]_{steady-state}^{channeled} \tag{13}$$

1.3.2.4 Transient versus steady state kinetics

For a two-enzyme reaction cascade, a steady-state can only be reached if $V_{max}^{(2)} > V_{max}^{(1)}$. Otherwise, the overall concentration of intermediate B will increase continuously until substrate A is completely consumed. Under these conditions, however, the rate of product formation in the steady state will be limited by the first reaction, not the second enzyme, and thus be identical irrespective of whether intermediate channeling occurs or not. The identical rates are evident in the identical slopes for C in Figure 7 (green). Thus, steady-state kinetics are not suited for investigating cascade reactions for either $V_{max}^{(2)} > V_{max}^{(1)}$ or $V_{max}^{(2)} < V_{max}^{(1)}$.

Steady-state analysis can be used, however, with a scavenger enzyme E3 that competes for the intermediate (see Scheme 4). This approach was exploited in establishing evidence for oxaloacetate channeling between mitochondrial malate dehydrogenase and citrate synthase using different concentrations of the competing enzyme aspartate transaminase (AAT) (Scheme 5) [28]. Genetic fusion of malate DH and citrate synthase increased the rate of citrate synthesis in the presence of aspartate transaminase. The double negatively charged oxaloacetate is believed to be channeled via positively charged patches on the protein surface. This hypothesis is supported by the observation that channeling is abolished at high ionic strength.

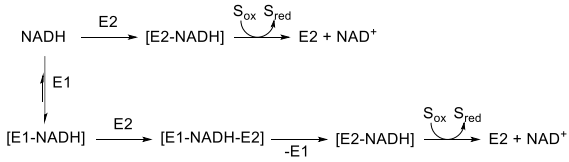


Scheme 5. Competition experiment between citrate synthase and aspartate transaminase for the intermediate oxaloacetate generated by malate dehydrogenase.

1.3.2.5 Enzyme buffering method

The enzyme buffering method involves “buffering” the concentration of intermediate B by binding to an enzyme. It has been primarily used to investigate transfer of NADH between alcohol dehydrogenases, particularly those involved in mitochondrial redox processes, which are presumed to bind one another weakly [29]. Conditions are chosen such that the concentration of E1 is higher than that of NADH, resulting in the majority of NADH being bound to E1. If [E2] is much smaller

than [E1] and the K_D for the E1-NADH complex, and if $k_{cat}^{(2)}$ and $K_m^{(2)}$ are known, the reaction velocity of E2 can be estimated by plugging the calculated concentration of free NADH into the Michaelis-Menten equation (Scheme 6). If the observed rates in such a system are higher than the predicted rates, one can infer that NADH is directly transferred from E1-NADH to E2 [30–32]. The viability of such an analysis has been challenged, however [33].



Scheme 6. The enzyme buffering method was used to investigate direct channeling of NADH between alcohol dehydrogenases. Figure adapted from reference [21].

1.3.2.6 Modeling cluster-mediated channeling by reaction-diffusion equations

Douglas and coworkers developed a reaction-diffusion mathematical model to rationalize the observed changes in reaction rates observed upon co-encapsulating two sequentially acting enzymes in a virus-like particle [34]. In this model, which combines concentration-dependent enzyme reaction rates and Fick's second law of diffusion, the intermediate is produced at the center of a sphere and consumed on its surface. In order to solve the heat transfer equation that describes the spatial distribution of the intermediate as a function of its rate of formation and consumption, the authors had to introduce two boundary conditions that deviate from physical reality: i) the temporal evolution of intermediate concentration at the point of its generation is solely described by the rate of its formation (k*t) and thus fails to account for its reduction by diffusion; and ii) a no-flux boundary at the surface of the sphere corresponds to complete consumption by the second enzyme at this point in space. Nevertheless, the conditions that this model predicts to favor substrate channeling are consistent with the generally accepted findings of McCammon and co-workers (see section 1.3.2.3, page 30) [22,27]. In particular, the maximum rate of the second enzyme has to be higher than the rate at which the intermediate is formed, which can be enforced by decreasing the rate of enzyme 1, increasing k_{cat} of enzyme 2, and/or reducing its K_m value.

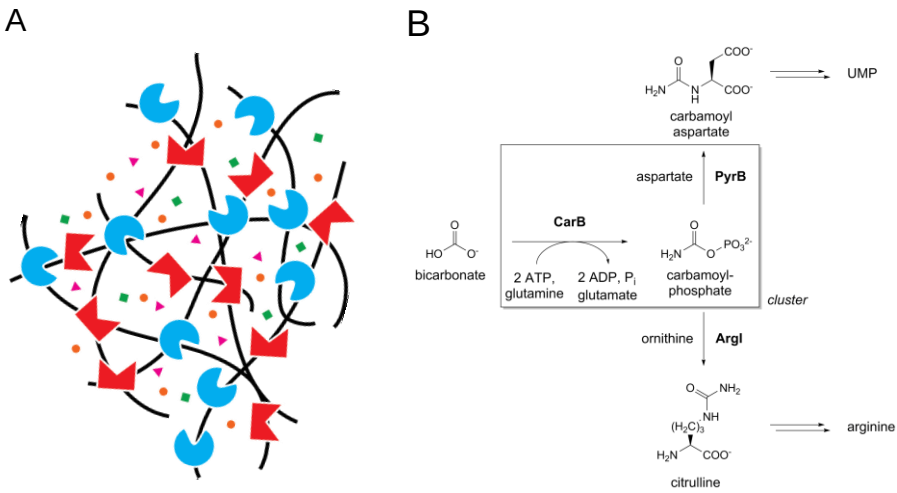


Figure 8. Metabolic channeling by enzyme clustering. A) Scheme of densely packed enzymes attached to the structural scaffold forming a multi-enzyme cluster, or metabolon. B) Branching of carbamoyl phosphate was employed as a model system to investigate the effect of enzyme clustering on substrate channeling. Figure adapted from reference [35].

The effect of enzyme clustering on metabolic channeling, for example by encapsulation in a protein container, has been analyzed by Wingreen and co-workers [35]. In an elaborate mathematical model, they determined the optimal spatial distribution of two sequentially acting enzymes in terms of pathway efficiency, defined by the rate of product synthesis divided by the rate at which substrate is provided. The difference between the two rates arises from substrate decomposition, processing by an alternative pathway, or escape from the enzyme cluster. The authors invoke basins at the center of which enzymes are clustered within a sphere, approximating the situation within a living cell where multiple clusters might be equally distributed throughout the cytosol. Using kinetic constants for carbamoyl phosphate synthetase (CarB, an enzyme that itself channels its substrate through a tunnel, see section 1.4, page 36) and aspartate carbamoyl transferase (PyrB) (Figure 8B), a decay rate of 10 s^{-1} was calculated for the intermediate carbamoyl phosphate. In the optimal configuration, both enzymes were co-clustered in a sphere with r = 0.26 µm and individual clusters were separated by ~6.5 µm. The spacing between clusters has an optimal value for the spacing, because diffusion of the substrate to the enzymes becomes limiting if the enzyme clusters are too sparsely distributed within the cytosol. Conversely, if the spacing between clusters is too narrow, smaller enzyme agglomerates result, increasing the likelihood that the intermediate escapes the

cluster before being processed. Ideally, the enzymes should be distributed uniformly within a cluster and packed as densely as possible (Figure 8A). In such a situation, pathway efficiency is predicted to be about six-fold higher than if the enzymes were freely diffusing in the cytosol. Below the optimal distance between clusters, overall efficiency decreases substantially with no advantage expected for agglomerates closer than 1 µm.

To test their predictions experimentally, the authors studied the consumption of carbamoyl phosphate at the branch point between arginine and uracil biosynthesis, where PyrB catalyzes the committed step towards UMP synthesis and ornithine carbamoyl transferase (ArgI) the dedicated step for arginine synthesis (Figure 8B). Although simply fusing CarB and PyrB did not induce channeling, agglomeration resulting from overexpressing the fusion protein in *E. coli* did. The cells became pseudo-auxotrophic for arginine. The concentrations of carbamoyl aspartate and its downstream metabolites increased by as much as 50-fold, accompanied by a drop in both citrulline and arginine pools and a buildup of their precursors ornithine and acetyl-glutamate [35].

1.4 Enzyme co-localization in nature

1.4.1 Substrate channeling in stable multi-enzyme complexes

Picture BASF's Verbund site in Ludwigshafen, Germany, the world's largest chemical production area. Core petrochemicals are produced in approximately 200 plants. The 36,000 people operating these plants direct raw materials, intermediates and products between single reactors, resorting to a 2,800 km network of pipelines. Now imagine all these reactions taking place in a single large reactor, all at the same time, and you do not come anywhere close to the chemical complexity within a living cell. Intricate regulatory mechanisms at the transcriptional, translational and protein level not only maintain the concentrations of essential metabolites, but also allow the chemical machinery to respond to external stimuli, such as changes in temperature, pH, ionic strength and the availability of nutrients. However, only a handful of metabolites are of global cellular importance. These include the energy-rich nucleoside triphosphates, redox agents like NADH and $FADH_2$, one-, two- and three-carbon equivalents such as methylene-THF, acetyl-CoA and pyruvate, as well as some core anabolic building blocks produced in the pentose phosphate pathway or the TCA cycle. The majority of molecules, however, are substrates for only one or a few enzymes. As a consequence, their scattered distribution over the entire cytosol would be unnecessary and inefficient. It is therefore no surprise that enzymes operating within the same pathway are often co-localized. The cells of higher organisms designate specific tasks to subcellular spaces such as mitochondria for

catabolism and energy production, chloroplasts for photosynthesis, or peroxysomes for fatty acid degradation. In these organelles, limited diffusion through the outer boundary, a phospholipid bilayer, increases the local concentrations of metabolites. On a lower level of organization, also found in prokaryotes, sequentially acting enzymes can form multi-enzyme complexes. In such structures, high local concentrations of intermediates are achieved by three principal mechanisms: i) the metabolite is covalently tethered to the complex; ii) supramolecular interactions on internal or external surfaces of proteins (or protein complexes) guide metabolites from one active site to the next; or iii) small-molecule diffusion is limited by a dense network of macromolecules and high local concentrations of enzymes consume the metabolite before it can leave the realm of the microdomain (Figure 9).

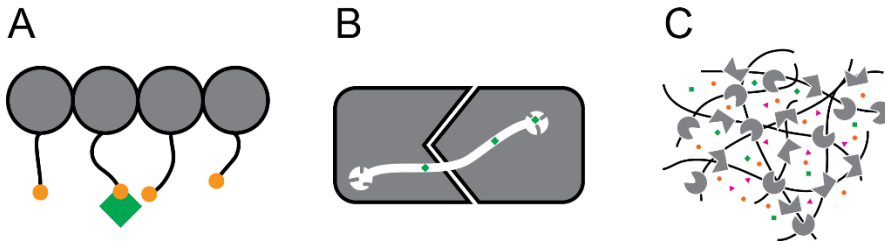
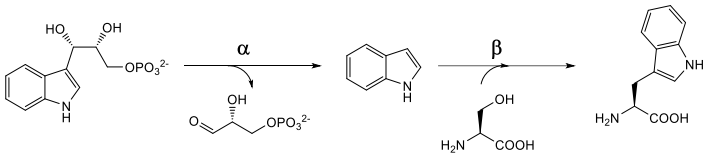


Figure 9. Three types of substrate channeling commonly encountered in nature. A) The substrate is covalently tethered to an enzyme complex and transferred between catalytic domains by hand-off to another tether. B) Multiple active sites of an enzyme are connected by an internal tunnel or a surface channel confining the intermediate by supramolecular interactions. C) Metabolons, microdomains of enzymes involved in a metabolic pathway, are densely clustered, potentially mediated by the cytoskeleton. The high enzyme density increases the likelihood of an intermediate being processed by an enzyme before diffusing out of the cluster. The size of a metabolon may span > 100 nm.

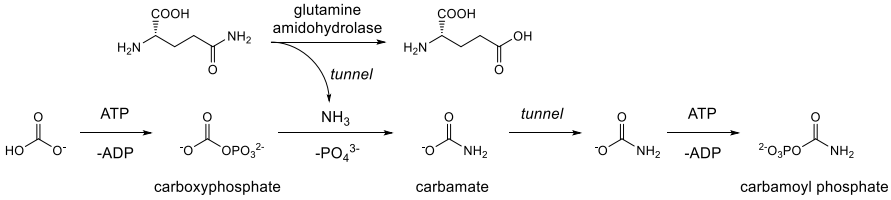
A well characterized example for the first mechanism is mammalian fatty acid synthase in which the nascent natural product is covalently attached to the enzyme complex via a phosphopantetheine prosthetic group and iteratively extended by repeated visits to a series of functionally distinct active sites [36]. Although the substrate remains bound to the same cofactor in fatty acid synthase, in the pyruvate dehydrogenase complex the metabolite is transferred between two active sites by handover as a covalent adduct via two prosthetic cofactors, thiamine pyrophosphate and lipoamide. Remarkably, the cores of the pyruvate dehydrogenase complex exhibits the same principles of quasi-equivalency as viral shells (see chapter 3). They are comprised of either a 24-meric cube (prokaryotes) or a 60-meric pentagonal dodecahedron (eukaryotes) of dihydrolipoyl transacetylase, which serves as a scaffold binding several copies of pyruvate decarboxylase and dihydrolipoamide dehydrogenase [37,38]. The concept of transferring covalently bound intermediates

between multiple prosthetic groups in large enzyme complexes is not only encountered in primary metabolism, but is also a hallmark of the syntheses of secondary metabolites such as polyketides and non-ribosomal peptides [39].

Direct transition of a metabolite between two active sites of an enzyme or enzyme complex without transient release into bulk solvent is referred to as substrate channeling [40]. The most prominent example is the heterotetrameric $\alpha_2\beta_2$ tryptophan synthase complex in which a tunnel connects the active sites of the α and β subunits [41,42]. Indole, formed in a retro-aldol reaction at the α site, diffuses to the β site at a rate >1000 s^{-1}, where it is added to the side-chain of serine in a PLP-catalyzed reaction (Scheme 7).



Scheme 7. Reaction sequence catalyzed by tryptophan synthase.



Scheme 8. Carbamoyl phosphate contains three active sites between which the reactive intermediates ammonia and carbamate are channeled via tunnels through the protein.

Carbamoyl phosphate synthase comprises three active sites that are interconnected by a 96 Å long tunnel [43]. Ammonia generated by hydrolysis of glutamine to glutamic acid is channeled to the site of carboxyphosphate formation, with which it reacts to form carbamate. Carbamate, which has a half-life of ~70 ms at pH 8.0 [44], is then channeled to a second phosphorylation site, where carbamoyl phosphate, a building block in pyrimidine biosynthesis, is formed (Scheme 8). Similar ammonia channels are postulated for guanosine monophosphate synthetase, glutamine phosphoribosylpyrophosphate amidotransferase and asparagine synthetase [40]. All of these enzymes utilize ammonia generated from hydrolysis of glutamine and activate their substrate for aminolysis by phosphorylation.

Thymidylate synthase (TS) catalyzes the transfer of a methyl group from 5,10-methylene tetrahydrofolate (5,10-CH_2-THF) to deoxyuridine monophosphate yielding deoxythymidine monophosphate and dihydrofolate (DHF). Reduction of DHF by dihydrofolate reductase (DHFR) produces tetrahydrofolate (THF), which is converted back into 5,10-CH_2-THF by serine hydroxymethyltransferase. In most organisms, TS and DHFR are expressed as separate polypeptides, but some parasitic protozoa produce both enzymes as a single polypeptide (DHFR-TS). Although kinetic experiments suggest channeling of DHF between the TS and DHFR active sites, there is no obvious tunnel between both domains [45–47]. Instead, a positively charged groove across the protein surface was identified, which likely guides negatively charged DHF (often bearing multiple γ-linked glutamyl moieties) from one active site to another (Figure 10) [48,49]. In humans, TS and DHFR are located on two different polypeptides and have not been isolated as a complex. Based on in silico docking studies, Wang and McCammon suggest that a similar electrostatic channel might also be formed in a non-covalent human DHFR-TS complex [50].

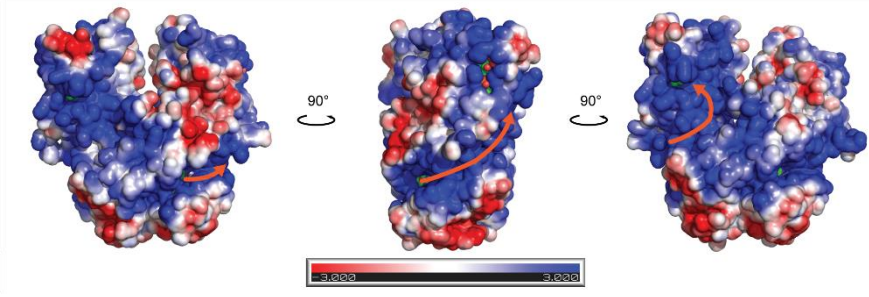


Figure 10. Electrostatic highway on the surface of the *Plasmodium falciparum* DHFR-TS homodimer [51]. The orange arrow indicates the positively charged channel on the protein surface connecting the transmethylation active site to the DHF binding pocket of the hydride transfer active site. The electrostatic surface potential of PDB structure 1J3I [52] was calculated using the APBS webserver (pH 7.0, I = 150 mM) [53] and visualized using PyMOL [54]. The color gradient indicates the local electrostatic surface potential between ± 3 k_BT/e.

1.4.2 Metabolons in higher organisms

The existence of large, weakly associated multi-enzyme complexes, so called metabolons [15,55], in higher organisms is a matter of intense debate. Such aggregates are potentially dynamic, characterized by weak protein-protein interactions and high dissociation rates, making isolation difficult if not impossible. Within cells, macromolecular crowding, high viscosity, and a difficult to penetrate cytoskeletal network might lead to increased local concentrations of metabolites and enzymes.

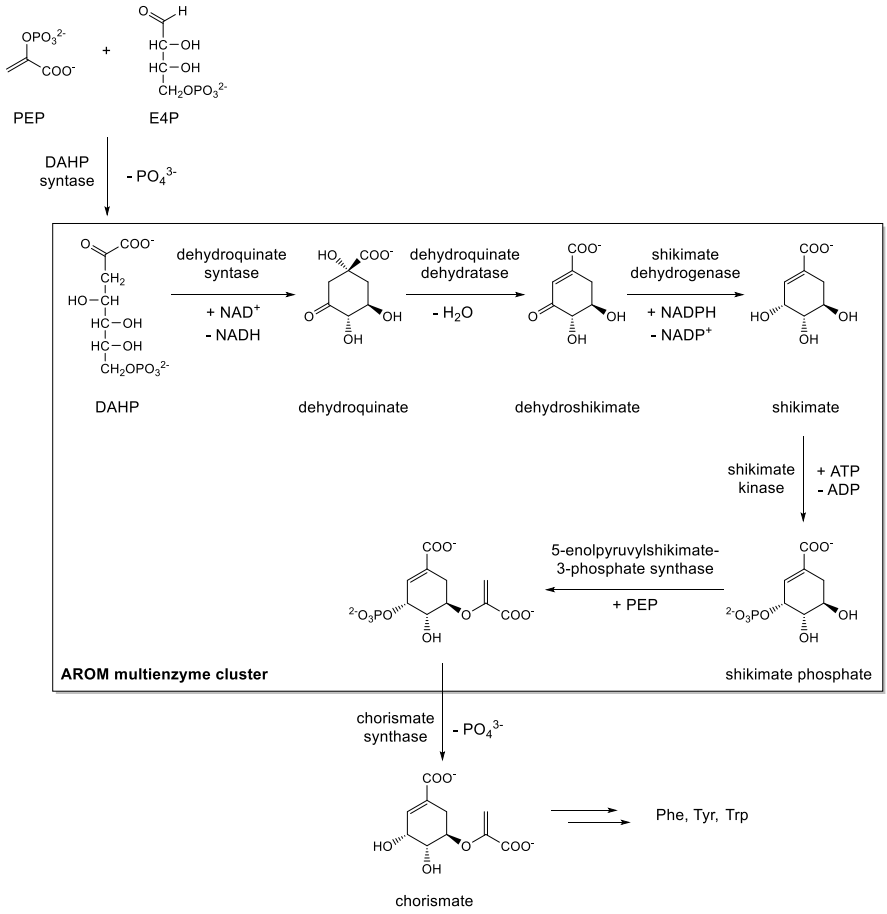
By obstructing diffusion of the catalysts from the supramolecular complex, such crowding effects could potentially compensate for otherwise weak protein-protein interactions by increasing local protein concentration. In eukaryotes, metabolon assembly might be regulated by elements of the cytoskeleton, adding another difficulty to its in vitro reconstitution [56].

Metabolite channeling in the primary and secondary metabolism of plants has been reviewed [57–59]. Multi-enzyme complexes comprised of several Calvin cycle enzymes have been isolated from spinach chloroplasts and are likely localized to the thylakoid membranes [60]. Particular attention has been paid on the regulation of the shikimate and phenylpropanoid pathways, which constitute important metabolic branch points and provide central precursors for aromatic plant natural products such as flavonols, anthocyanins, coumarins, tannins and lignin [59,61].

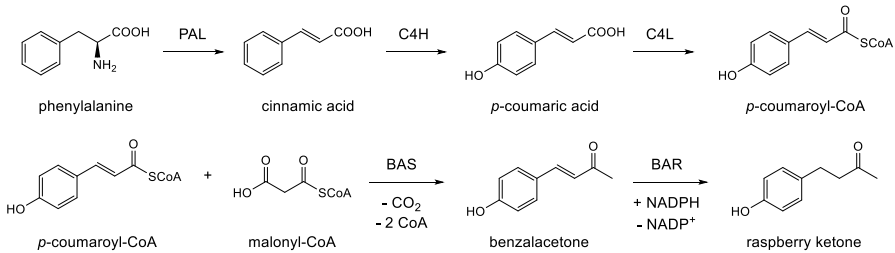
In fungi and higher plants, enzymes of the shikimate pathway form a so-called AROM multienzyme complex (Scheme 9) [62]. Five consecutively acting enzymes of the polyaromatic branch are even expressed as a single polypeptide in some fungi [63]. The main role of the AROM complex might be to regulate pathway flux since many organisms possessing this enzyme complex can utilize intermediates of the polyaromatic pathway as their sole carbon source and therefore require a mechanism to bring the anabolic reactions to a halt. Binding of the first substrate, 3-deoxy-D-arabino-heptulosonate 7-phosphate (DAHP), to the enzyme complex diminishes its proteolytic degradation, presumably through increased population of conformations that do not expose protease cleavage sites [64]. Additionally, binding of DAHP increases k_{cat} and/or decreases K_m values of all enzymes except shikimate kinase, making it rate-limiting and thus rendering the pathway flux susceptible to the availability of ATP and product inhibition exerted by ADP [65].

The first steps of the phenylpropanoid pathway is presumably localized in a membrane associated metabolon (Scheme 10). Dixon and co-workers found evidence for substrate channeling of cinnamic acid between phenylalanine amino-lyase (PAL) and cinnamic acid 4-hydroxylase (C4H) in isotope-dilution experiments [66]. Association of PAL to membrane bound C4H was observed using a PAL-GFP fusion protein [67]. Additional photobleaching studies suggest that the underlying protein-protein interactions are weak. Formation of membrane bound enzyme complexes is also observed for several downstream branches of the phenylpropanoid pathway leading to anthocyanins or lignin [59]. Given the lack of substrate specificity that many of the involved enzymes display in vitro, metabolon formation might be a way to control metabolic flux. Besides reducing the dimensionality of diffusion by providing a barrier surface (see section 1.5.3, page 52) it has been pointed out that membrane microdomains provide a solvent microenvironment

that allows accumulation of lipophilic metabolites above their solubility in aqueous solution (Figure 11) [59].



Scheme 9. Shikimate pathway for biosynthesis of aromatic amino acids initiated by formation of 3-deoxy-D-arabino-heptulosonate 7-phosphate (DAHP) from phosphoenol pyruvate (PEP) and erythrose-4-phosphate (E4P) by DAHP synthase. In some organisms, five enzymes form a multienzyme complex (AROM complex), which allows regulation of the entire pathway by the availability of ATP [62].



Scheme 10. Phenylpropanoid pathway leading to raspberry ketone. Cinnamic acid formed by phenylalanine ammonia-lyase (PAL) is channeled to a membrane-bound P450 hydroxylase (C4H) introducing a hydroxy group in para-position [66,67]. *p*-coumaroyl-CoA formed by *p*-coumarate-CoA ligase (C4L) is a precursor for manifold aromatic natural products such as raspberry ketone. The latter is obtained after decarboxylative Claisen condensation with malonyl-CoA catalyzed by benzalacetone synthase (BAS) [68,69] followed by enoate reduction by benzalacetone reductase (BAR) [70].

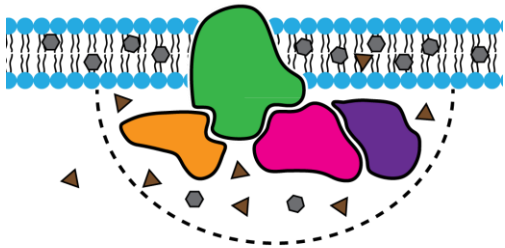
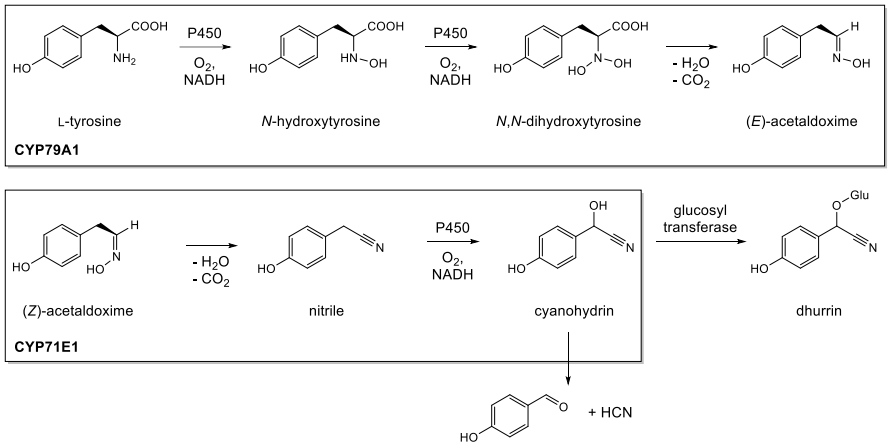


Figure 11. Membrane bound metabolon. The lipid bilayer reduces diffusion in one dimension and might enhance local concentration of lipophilic metabolites.

The benefit of metabolite channeling is immediately apparent in the biosynthesis of dhurrin, a cyanogenic glycoside produced by many plants as a repellent against herbivores (Scheme 11). A series of P450-catalyzed monooxygenations convert tyrosine to a cyanohydrin. If this metabolite escaped glycosylation, it would release toxic cyanide, which poisons plants and animals by incapacitating heme-containing proteins [71]. Using isotope-labeled tyrosine, *N*-hydroxy-tyrosine, aldoxime and nitrile derivatives, channeling was observed for *N*-hydroxy-tyrosine and the nitrile, but not for the aldoxime [72]. However, since it is now known that the steps converting tyrosine to aldoxime are all catalyzed by CYP79A1, whereas the steps converting aldoxime to cyanohydrin (detected as an aldehyde) are all catalyzed by another enzyme CYP71E1 [73], these results rather reflect the ability of a single active site to catalyze multiple reactions than actual substrate channeling between enzyme active sites. Despite lack of kinetic evidence, cyanohydrin seems to be channeled between CYP71E1 and the glucosyltransferase. Membrane-bound enzymes

CYP79A1 and CYP71E1 were able to localize a soluble yellow fluorescent protein-glucosyltransferase fusion protein to the membrane of the endoplasmatic reticulum, suggesting formation of a metabolon [74]. No dhurrin was detected in transgenic plants in which the interaction between the CYPs and the glucosyltransferase was disrupted. The indispensability of such a multienzyme complex for dhurrin formation suggests that proximity of the site of cyanohydrin generation and consumption is crucial. Interestingly, the entire metabolon could be heterologously expressed in *Arabidopsis thaliana*, enabling a plant originally devoid of this pathway to accumulate dhurrin to 4 % dry weight with marginal pleiotropic effects [75].



Scheme 11. Biosynthesis of dhurrin necessitates direct transfer of a cyanohydrin intermediate to avoid release of toxic cyanide. Scheme adapted from reference [59].

Two topical examples for which evidence of metabolon formation in mammals has been gathered are purinosomes in HeLA cells and glycolysis in erythrocytes. Benkovic and coworkers discovered that six enzymes with a total of ten catalytic sites necessary for de novo synthesis of inosine monophosphate, the direct precursor to the nucleobases AMP and GMP, reversibly co-localize into so-called purinosomes when HeLa cells were cultured on purine-depleted medium [76]. While concerns were raised that co-localization of these enzymes might be an artifact resulting from nonspecific aggregation of the overexpressed GFP-fusion proteins under stress conditions [77], Benkovic's group demonstrated that microtubule assembly is prerequisite for purinosome formation [56]. In an isotope tracer experiment with ^{14}C-labeled glycine, a biosynthetic purine precursor, a 42 % increase in ^{14}C-incorporation into cellular purines was observed under purine-depleted versus purine-rich

conditions. Addition of a microtubule-disrupting compound suppressed incorporation by 36 % under purine starvation, whereas no significant effect was observed when purines were abundant.

Interactions between the enzymes involved in glycolysis and the cytoskeleton, in particular F-actin, have sparked speculation about a glycolytic metabolon. However, despite decades of research, no direct evidence for such an enzyme supercomplex has been presented, possibly due to the low affinity of individual protein-protein interactions and the resulting dynamic nature of the putative glycolytic metabolon [78,79]. Assembly of a multienzyme complex comprising several glycolytic enzymes at the cellular membrane has been observed in human erythrocytes. The N-terminus of the anion channel protein band 3 has been identified as the binding motif for glyceraldehyde-3-phosphate dehydrogenase, fructose 1,6-bisphosphate aldolase and phosphofructokinase [80]. Later, a second site in this channel protein was found to bind pyruvate kinase and lactate dehydrogenase as well as glyceraldehyde-3-phosphate dehydrogenase and fructose 1,6-bisphosphate aldolase [81]. All glycolytic enzymes are displaced from the membrane upon tyrosine phosphorylation of band 3 or binding of deoxy hemoglobin. Lacking mitochondria, erythrocytes rely entirely on anaerobic glycolysis to produce ATP, which is primarily consumed by transmembrane ion pumps. From a metabolic point of view, co-localizing sites of ATP production and consumption is reasonable. Indeed, experiments using a fluorescent ATP analog point to the existence of localized ATP-pools entrapped by the the cell membrane and cytoskeletal ankyrin and β-spectrin proteins. Covalent protein labeling using a photoactivatable ATP analog places band 3, glyceraldehyde-3-phosphate dehydrogenase and, by extrapolation, the putative glycolytic enzyme complex in the vicinity of ATP-dependent Na^+/K^+ and Ca^{2+} pumps [82].

1.5 In vitro enzyme co-localization

1.5.1 Polymersomes

Natural multienzyme complexes have inspired chemists to create and test analogous artificial constructs, not least to highlight the utility of novel methods for supramolecular assembly. For example, polymersomes are vesicles assembled from amphiphilic block copolymers that resemble liposomes in shape, size and topology but are biorthogonal and considerably more resilient towards mechanical stress [83,84]. The constituent polymers can be chemically modified at will resulting in tailored polymersomes that can respond to a variety of external stimuli, rendering them sought-after vehicles for drug delivery [85]. The ability to create nested polymersomes, i.e. smaller vesicles within larger vesicles, led to their utilization as artificial cells with tailored subcellular compartments containing catalytic entities. In principle, enzymes can be localized in the aqueous lumen, within the amphiphilic

layer or on its surface. Exploitation of this setup for the creation of biomimetic nanoreactors and artificial organelles has been reviewed [86–88].

Van Hest and co-workers developed a procedure to target enzymes either to the hydrophilic core or the hydrophobic membrane of a styrene/isocyanoalanine diblock copolymer-based polymersome [89]. In analogy to liposomes, these polymersomes are hollow spheres with a shell assembled from amphiphilic polymers. The authors carried out a three-enzyme cascade reaction, using exogenous *Candida antarctica* lipase B (CalB) to deacetylate tetra-acetylated β-glucopyranose, which then penetrated the polymersomes and got oxidized to the lactone by oxygen-dependent glucose oxidase (GOx) residing in the hydrophilic lumen (Figure 12). The side product of the latter reaction, H_2O_2, was detected spectrophotometrically through oxidation of chromogenic 2,2'-azino-bis(3-ethylbenzthiazoline-6-sulphonic acid) (ABTS) by horseradish peroxidase (HRP) located in the membrane. While the control reaction with free enzymes stopped abruptly after 170 min at 0.8 % conversion, the three-enzyme-polymersome system proceeded for 25 hours to give approximately 20 % conversion.

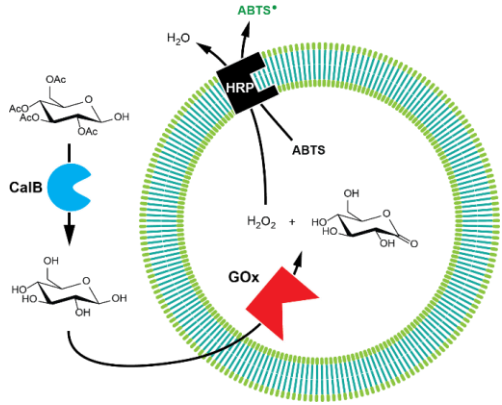
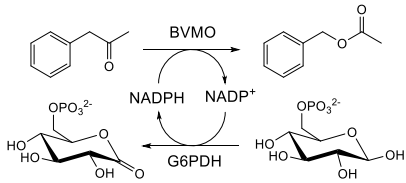


Figure 12. Polymersome based cascade reaction. A vesicle formed from diblock copolymers (polystyrene in teal, polyisocyanide in green) topologically separates three enzymes: CalB lipase in the bulk solvent, glucose oxidase (GOx) enclosed within the polymersome and horse radish peroxidase (HRP) associated with the polymer-bilayer. CalB deacylates acylated glucose, which can then be oxidized to glucono-*δ*-lactone by GOx, the side product H_2O_2 being used to oxidize chromogenic ABTS by HRP. Figure adapted from reference [89].

The improved activity of enzymes in polymersomes is often attributed to protection from proteases. Deliberate addition of subtilisin A to a mixture of free GOx and HRP led to complete abolishment of enzymatic activity, whereas activity was retained when GOx and HRP were individually localized within polymersomes [90].

Thus, protection from proteolysis provides one possible explanation for why enzymes contained in polymersomes often retain high activity over weeks or months, whereas they last several days at most when free in solution [89–91].

The same styrene/isocyanoalanine block copolymer was also used to create a polymersome co-factor regeneration system for an NADPH-dependent Baeyer-Villiger monooxygenase. To that end, glucose-6-phosphate dehydrogenase was enclosed within the polymersome and the monoxygenase immobilized on its surface (Scheme 12). However, this system displayed reduced substrate conversion rates compared to the enzymes free in solution [92].



Scheme 12. Design of a co-factor regeneration system. A Baeyer-Villiger monooxygenase (BVMO) utilizes NADPH as an essential cofactor that can be regenerated by co-encapsulation of glucose-6-phosphate dehydrogenase (G6PDH).

A system for quenching reactive oxygen species in vivo has been reported by Palivan and coworkers [91]. A superoxide dismutase (SOD) and a peroxidase were co-encapsulated in a polymersome assembled from triblock copolymers (Figure 13). Addition of a non-specific transport channel protein to the polymer-enzyme mixture rendered the resulting polymersome permeable to small molecules such as the fluorogenic Amplex Red used to detect peroxidase activity. Only when the transport channel protein was in place, human cells treated with these nanocontainers showed localized fluorescence from the Amplex Red peroxidation product resorufin upon incubation with a sensitizer generating superoxide. This system convincingly demonstrates the utility of cascade processes for detoxification of superoxide to water and oxygen via a hydrogen peroxide intermediate.

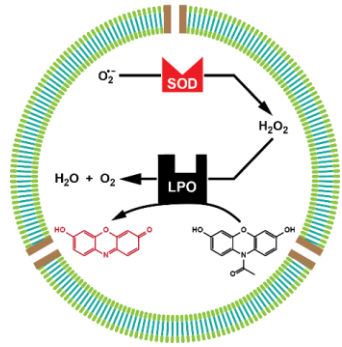


Figure 13. Polymersome with embedded channel proteins (brown) for superoxide detoxification using encapsulated superoxide dismutase (SOD) and a peroxidase (LPO). Peroxidase activity is detected by oxidizing the fluorogenic dye Amplex Red (black) to resorufin (red). Figure adapted from reference [91].

The groups of Lecommandoux and van Hest constructed a hierarchical polymersomal cell mimic that co-confined three different enzymes: a Baeyer-Villiger monooxygenase, an esterase, and an alcohol dehydrogenase (Figure 14) [93]. The individual enzymes were encapsulated in polymersomes with diameters in the submicrometer regime, which were then confined within larger polymersomes. The resulting "organelle-containing cell mimics" had diameters in the double-digit micrometer range (Figure 14). When the esterase was substituted by a peptidase, the confinement of the latter within a small polymersome led to a significant increase in the overall reaction rate. The authors rationalized this observation by arguing that encapsulation of the protease within a polymersome slows down or altogether prevents degradation of the other enzymes in the pathway.

Städler and co-workers constructed another variation on higher-order, polymer-based structures, which they named capsosomes. In one example, multiple different liposomes were co-confined within a polymeric capsule with the ultimate goal of creating a sensor for uric acid (Figure 15) [94]. Uric acid can be detected by a coupled enzyme assay in which the hydrogen peroxide formed in its uricase-catalyzed oxidation is used to oxidize a fluorogenic substrate. In biological samples, the second peroxidation step is often impeded by the presence of reducing ascorbic acid. Capsosomes with an approximate diameter of 5 µm composed of equal amounts of liposomes containing either uricase, horseradish peroxidase or ascorbate oxidase yielded about three-fold higher fluorescence upon incubation with a sample containing uric acid, ascorbic acid and the fluorogenic reporter Amplex Ultra Red as compared to a similar capsosome lacking ascorbate oxidase. For this demonstration, however, it was irrelevant whether co-confinement of ascorbate oxidase provided a kinetic advantage and was thus not tested.

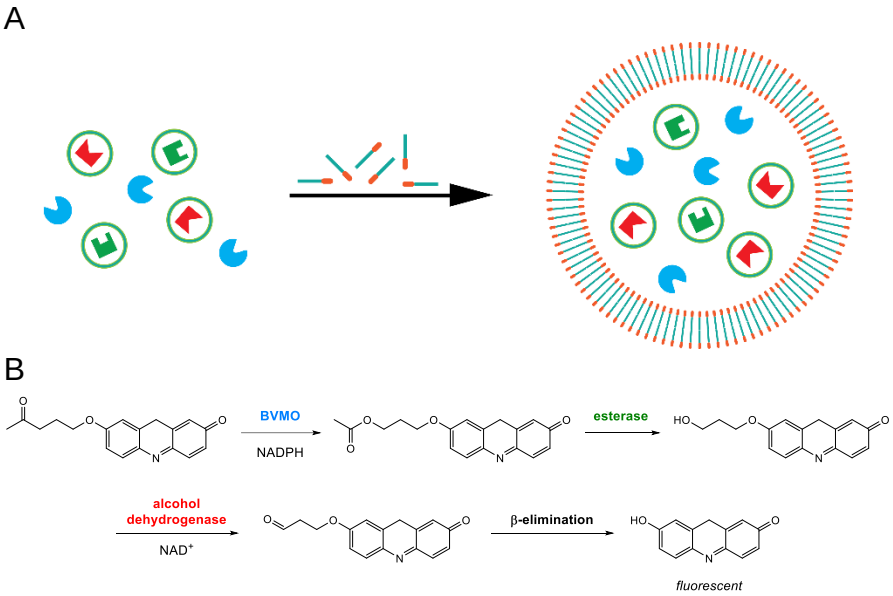


Figure 14. A polymersome organelle-containing cell mimic. A) Individual enzymes encapsulated in submicrometric polymersomes assembled from styrene/isocyanoalanine diblock copolymers are mixed with free enzymes, reaction components, and styrene/ethylene oxide diblock copolymers to form polymersomes with diameters >1 µm. The polymersomes-within-polymersomes are supposed to mimic organelles within cells. B) The reaction cascade catalyzed by three encapsulated enzymes—a Baeyer-Villiger monooxygenase (BVMO), an esterase, and an alcohol dehydrogenase—ultimately yields fluorescent resorufin. Figure adapted from reference [93].

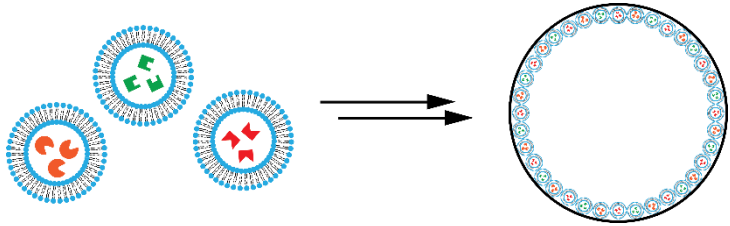


Figure 15. A capsosome consisting of three different kinds of liposomes (blue) containing either uricase, horseraddish peroxidase or ascorbate oxidase enclosed in a polydopamine shell (black). This assembly is used for improved uric acid detection in biological samples. Figure adapted from reference [94].

Mann and coworkers created proteinogenic liposome analogs, so called proteinosomes, in which the amphoteric moiety is made up from hydrophilic proteins conjugated to hydrophobic polymers [95]. The enzymes were cationized (i.e. the

surface exposed carboxylates were reacted with 1,6-diaminohexane) and subsequently derivatized with hydrophobic polymers containing an activated ester [96]. In a water-oil dispersion, mixtures of the enzyme-polymer conjugates formed water-containing vesicles that were stabilized by covalent cross-linking (Figure 16). The resulting micro-compartments consisted of closely packed monolayers of conjugated protein-polymer building blocks, for example a mixture of the sequentially acting enzymes glucoamylase, glucose oxidase, and horseradish peroxidase. In light of the non-native conditions used during formation of the multi-enzyme proteinosomes, the three-order of magnitude reduction in overall reaction rate compared to free enzymes is not surprising. Interestingly, the reaction rate of a mixture of proteinosomes containing only one enzyme each was reduced 23 % compared to proteinosomes formed from all three catalysts. The authors attributed this difference to reduced diffusional constraints for the transfer of intermediates in the reaction cascade when glucoamylase, glucose oxidase and horseradish peroxidase are co-localized on the surface of the same compartment.

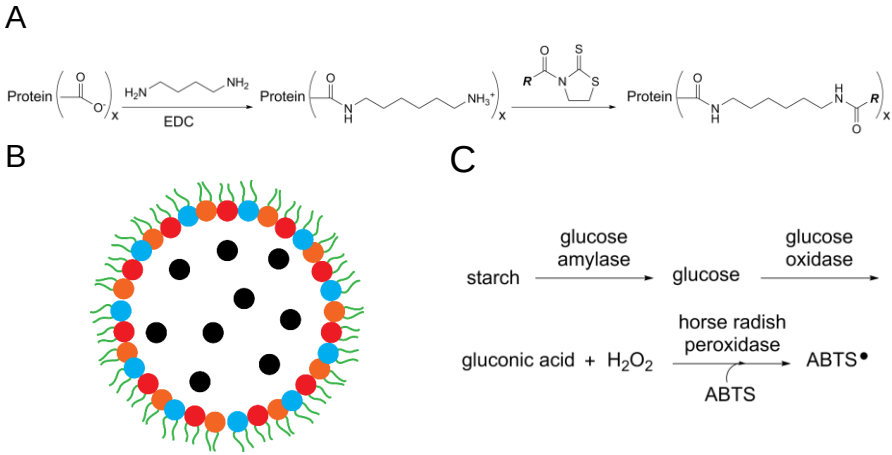
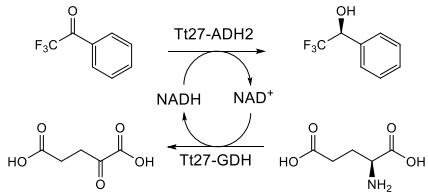


Figure 16. Proteinosomes formed from proteins modified with lipophilic polymers. A) Chemical derivatization of the proteins of interest by first modifying surface carboxylates with di-amines upon activation by EDC (cationization) before decorating them with lipophilic polymers *R*. B) Mixing glucose amylase, glucose oxidase and horse radish peroxidase (represented by differently colored spheres), conjugated with lipophilic polymers (green), with organic solvent induces formation of proteinosomes, i.e. vesicles in which the amphiphiles are polymer protein conjugates. Addition of nonderivatized proteins (black) localizes them in the lumen of the proteinosome. C) Scheme of the cascade reaction investigated using proteinosomes.

To date, there is no example reported in which a polymersome-based nanoreactor encapsulating multiple enzymes confers a clear kinetic advantage. The concentration of co-confined catalysts by passive encapsulation into polymer-encompassed spaces is probably too low to provide a crowding effect relative to bulk solution. Even if there was an effect, the experimental conditions are often not optimized towards detection of subtle enhancements. Consequently, Marguet et al. concluded that "enzymatic activity in polymersomes is often maintained longer than in free solution, certainly due to a shielding of their degradation by microbes and enzymatic proteolysis that constitutes one of the major and most concrete advantages of nanoreactors" [86]. Nevertheless, the cascade reactions reported in the polymersome literature tend to involve a restricted set of commercially available enzymes and substrates, presumably because chemical modification of enzymes typically requires large quantities of protein and most experimentalists have a limited inclination to produce or engineer alternative biocatalysts. Interdisciplinary collaborations between polymer chemists and enzymologists might present an opportunity to expand the range of polymersome-based biocatalytic cascade reactions.

1.5.2 Co-immobilization of biocatalysts

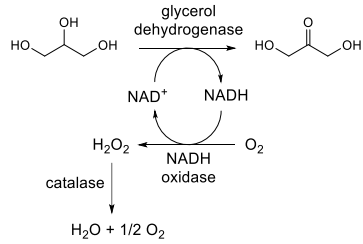
López-Gallego and coworkers compared different strategies for co-immobilization of three alcohol dehydrogenases and different cofactor regeneration systems on glyoxyl-agarose supports [97].



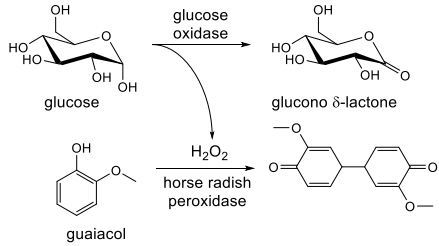
Scheme 13. Cascade reaction catalyzed by co-immobilized alcohol dehydrogenase (ADH) and glutamate dehydrogenase (GDH) from *Thermus thermophilus* on porous functionalized agarose beads.

When cofactor concentrations were limiting, the cofactor-recycling frequency (defined as rate of product formation per cofactor) improved 15.3-fold when an alcohol dehydrogenase was immobilized on the same porous beads (d ≈ 100 µm) as the glutamate dehydrogenase used to regenerate NADH versus when both enzymes were attached to different beads (Scheme 13). The higher concentration of cofactors in the vicinity of the active sites was further demonstrated by an approximately two-fold reduction in the apparent K_m value for NADH compared to the free soluble enzymes. Apparent co-factor affinity and recycling frequency could be improved by

another two-fold by using immobilization conditions that favor homogeneous spatial distributions of both enzymes over the entire porous bead and optimizing the ratio of alcohol dehydrogenase to glutamate dehydrogenase. Using a similar approach, the same research group improved the yield of 1,3-dihydroxy acetone from glycerol using an engineered glycerol dehydrogenase, an NADH oxidase for co-factor regeneration, and a catalase to remove the H_2O_2 byproduct, which would otherwise oxidize 1,3-dihydroxy acetone to glycolic acid (Scheme 14) [98]. Co-immobilization on the same bead, followed by cross-linking (to avoid dissociation of enzyme oligomers), reduced the accumulated H_2O_2 from 19 to 2 μM and increased to final yield 2.5-fold compared to a mixture of beads decorated with only one kind of enzyme.



Scheme 14. Three-enzyme reaction system immobilized on agarose beads for increased 1,3-dihydroxy acetone yield.



Scheme 15. Two-enzyme reaction system immobilized on graphene oxide mediated by polyacrylic acid.

The k_{cat} value for oxidation of guaiacol by a co-immobilized mixture of glucose oxidase and horseradish peroxidase on graphene oxide sheets mediated by polyacrylic acid is about twofold higher than for the free enzymes, while the K_m value also increased twofold (Scheme 15) [99]. The differences were attributed to substrate channeling, which given the increased apparent K_m is questionable. Indeed,

the authors described the conclusions drawn from their kinetic studies as “rudimentary”.

1.5.3 DNA origami scaffolds

The field of DNA origami exploits the high specificity with which DNA strands complement each other to design three-dimensional structures with high spatial control on the nanoscale [100]. A study published by Hao Yan’s group in 2012 attracted particular attention and sparked a variety of related experiments. By virtue of DNA origami, a single glucose oxidase and horseradish peroxidase pair was placed at various distances on a 2-dimensional tile [18]. Although decreasing the inter-enzyme distance from 65 to 45 and 25 nm only resulted in moderate enhancements of the overall reaction rate, a sharp increase was observed upon further reducing the distance to 10 nm (Figure 17).

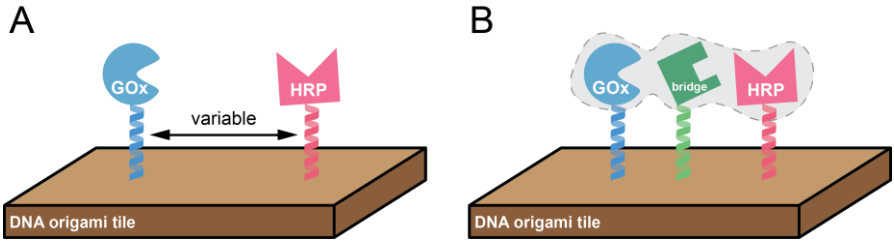


Figure 17. Glucose oxidase (GOx)/horseradish peroxidase (HRP) pair co-localized at various distances on a DNA origami tile. A) Schematic representation of the co-localization strategy. B) Scheme depicting how placing a non-catalytic protein between GOx and HRP can putatively promote substrate channeling by creating a continuous hydration shell (grey). Figures adapted from reference [18].

Employing a three-dimensional Brownian diffusion model, the authors excluded more efficient diffusional transfer of the H_2O_2 intermediate as the origin of the abruptly increased reaction rate, and rather attributed it to channeling through a shared hydration layer when the catalysts were placed within each other’s immediate vicinity. To corroborate this hypothesis, a non-catalytic protein was placed in between the two catalysts separated at 30 nm, which led to a 20 % increase in reaction rate in the case of a small protein (d ≈ 6 nm) and a 42 % increase for a larger one (d ≈ 16 nm). What is peculiar, however, is the lag of at least a minute for ABTS oxidation observed in all cases. A lag is generally expected for a cascade reaction because it reflects the time needed to build up sufficient concentrations of intermediate before the rate of its consumption (which depends on its concentration) equals the rate of production (see section 1.3.2.3, page 30). But if the intermediate were partially channeled by virtue of co-localization of two active sites, this effect

should be particularly important in the initial phase of the reaction during which differences in local concentration should be most strongly dependent on the distance between the sites of intermediate production and consumption. Yet, the lag appears very similar in all three cases. While it is conceivable that a surface channeling effect exists or that the active sites of the juxtapositioned catalysts become incredibly close, other mechanisms may equally well explain the observed phenomenon. Specifically, binding of either of the two protein catalysts to another protein might modulate their catalytic activity by, for example, changing their conformational energy landscape.

Computer simulations of Brownian dynamics in this experimental system suggest that the planar scaffold used to co-immobilize both enzymes enhances the probability of intermediate transfer by acting as a diffusive barrier and thus reduces the diffusional space from a sphere to a hemisphere [101]. According to these calculations, the probability of direct transfer of the intermediate is about three-fold higher when the two active sites directly face each other at a distance of 10 nm versus pointing in opposite directions. In an exemplary study, Wheeldon and coworkers demonstrated that the enhancement of horseradish peroxidase activity upon immobilization on a DNA scaffold depends on the binding energy of the substrate to DNA [26]. The maximum enhancement was observed for substrates that do not bind too weakly or too tightly to DNA. Since Yan and co-workers used DNA-origami scaffolds in all cases, this effect can be excluded as the origin of the sharp rate enhancement upon catalyst juxtaposition.

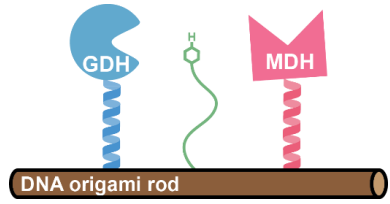


Figure 18. Glucose-6-phosphate dehydrogenase, malate dehydrogenase and NADH were covalently attached to a DNA scaffold in such a way that the NADH cofactor could transfer hydride equivalents between the two enzyme active sites. Figure adapted from reference [102].

Extending the concept of DNA origami, Yan and co-workers devised a nanoscaled reaction system in which a single NADH molecule was covalently attached to a DNA scaffold and placed in between two alcohol dehydrogenases, glucose-6-phosphate dehydrogenase (G6PDH) and malate dehydrogenase (MDH) (Figure 18) [102]. Covalent attachment of the cofactor increased the rate of reduction of ox-

aloacetate by glucose-6-phosphate more than 90-fold compared to the same concentration (0.1 μM) of freely diffusing NAD^+. Increasing the ratio of the rate-limiting MDH to G6PDH from 1:1 to 4:1, an additional 3-fold improvement in the overall rate was achieved. In an experiment in which malate dehydrogenase and lactate dehydrogenase competed for NADH, the MDH activity increased more than 30-fold when it was covalently attached to a scaffold tethered to G6PDH and NADH. In contrast, the activity of free lactate dehydrogenase dropped about 6-fold at the same time, highlighting the selectivity that such co-localization can provide.

1.5.4 Protein shells

In contrast to eukaryotes, bacteria and archaea lack subcellular compartments bounded by phospholipid bilayers. As a substitute, some organisms utilize polyhedral protein bodies that enclose a number of enzymes, presumably to create a chemical microenvironment distinct from the cytosol. These bodies, called bacterial microcompartments (BMCs) include the carboxysome, believed to increase the concentration of carbon dioxide at the active site of the carbon-fixating enzyme RuBisCO, the propanediol utilizing compartment (Pdu), and the ethanolamine utilizing compartment (Eut). The catabolic pathways catalyzed within Pdu and Eut involve propionaldehyde and acetaldehyde, respectively, and the alleged function of the compartments is containment of these cytotoxins [103]. Research on BMCs increased significantly over the last few decades, inspiring the use of proteinaceous shells from viruses and bacterial nano- or microcompartments to create artificial nanoreactors. As a result, there are manifold examples of enzymes localized to the lumen of protein containers using a variety of strategies (see chapter 4). The literature on cascade reactions within these capsids, however, is limited. The first and to our knowledge only published study on the effect of co-encapsulation in a protein container on the overall rate of a cascade reaction has been conducted in the Douglas laboratory [34]. A β-glucosidase (CelB), which catalyzes the hydrolysis of lactose to glucose and galactose, was co-encapsulated with a glucokinase (GLUK), which phosphorylates glucose to glucose-6-phosphate (G6P), in the capsid of bacteriophage P22 by covalently fusing both enzymes to the capsid scaffold protein (Figure 19).

Formation of glucose-6-phosphate was measured spectroscopically by monitoring the increase in NADPH concentration upon oxidation by exogenous glucose-6-phosphate dehydrogenase. Use of coupling enzymes to monitor cascade reactions in protein containers is only reasonable if the diffusion of the relevant molecules into and out of the capsid is not rate-limiting. Consequently, such systems can only detect substrate channeling arising from close proximity of both enzymes, but not from the container wall acting as a diffusion barrier. Previous studies by the Douglas

group demonstrated that the capsid wall of P22 does not impede diffusion of the substrates acetoin, 4-nitrophenyl-β-glucopyranoside, X-Gal or even NAD^+ [104,105].

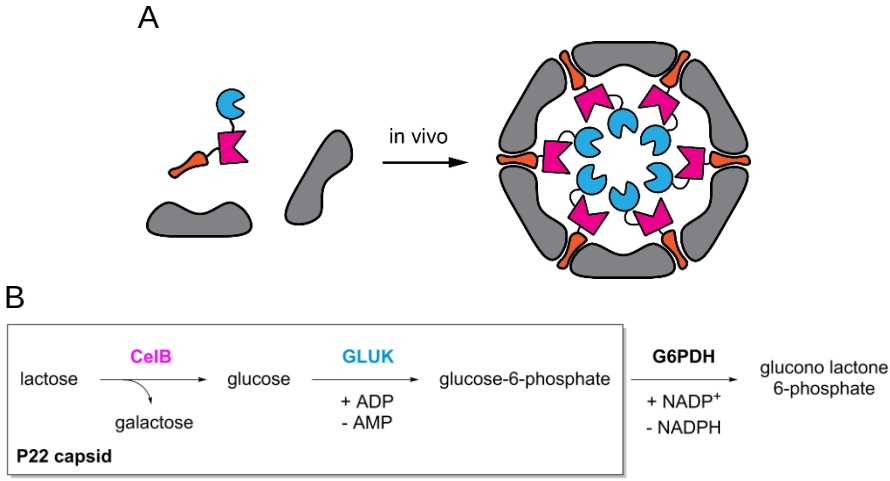


Figure 19. Cascade reaction within bacteriophage P22 capsids. A) Co-expression of the capsid coat protein (grey) and enzymes (magenta and blue) genetically fused to the capsid scaffold protein (orange) leads to incorporation of cargo enzymes into the assembled capsid. B) Cascade reaction investigated in the P22 capsid. Co-encapsulation of the glucosidase CelB and glucokinase (GLUK) transforms lactose to galactose and glucose-6-phosphate. The latter is detected spectrophotometrically using a coupled enzyme assay employing glucose-6-phosphate dehydrogenase (G6PDH) outside the capsid by following the increase in absorption upon NADPH formation [34].

Due to lack of an assay to determine lactose hydrolysis, CelB activity was determined with nitrophenyl-β-glucopyranoside and the observed trends were extrapolated to lactose. The key finding was that co-encapsulation yielded faster rates for the generation of glucose-6-phosphate if the first enzyme in the cascade, CelB, was slowed down by using a less efficient scaffold (GLUK-CelB-P22 instead of CelB-GLUK-P22) and by further inhibiting its activity with TRIS. The most likely explanation for the absence of an effect in the case of uninhibited CelB is that glucose concentrations are significantly higher than its K_m for its conversion by GLUK. In that case, GLUK would be saturated irrespective of whether it is co-encapsulated with CelB or not, and high local concentrations could not provide a kinetic advantage. Although not reported by Patterson et al., the kinetic parameters determined in their work allow indirect derivation of the steady state concentration of the glucose intermediate. During the initial lag phase, glucose is produced at a higher rate by CelB than it is consumed by GLUK. In the steady-state phase these rates must be equal (eq.

(14) and (15)) and, with knowledge of k_{cat} and K_m for GLUK, the steady state concentration of glucose can be determined. By virtue of covalent linkage, the concentration of CelB and GLUK monomers is always equal when encapsulated, which allows simplification of equation (15) to equation (16). Since the kinetic parameters of GLUK are known in all cases, and the rate of glucose consumption is identical to that of G6P production (eq. (17), the corresponding steady state concentrations of glucose *around* GLUK can be determined (eq. (18) and Table 1).

$$\frac{d[Glu]}{dt} = 0 = Formation - Consumption \tag{14}$$

$$\frac{d[Glu]}{dt} = 0 = \frac{k_{cat}^{CelB} \cdot [Lac] \cdot [CelB]}{K_M^{Lac} + [Lac]} - \frac{k_{cat}^{GLUK} \cdot [Glu] \cdot [GLUK]}{K_M^{Glu} + [Glu]} \tag{15}$$

$$\frac{d[Glu]}{dt} \cdot \frac{1}{[Enzyme]} = 0 = \frac{k_{cat}^{CelB} \cdot [Lac]}{K_M^{Lac} + [Lac]} - \frac{k_{cat}^{GLUK} \cdot [Glu]}{K_M^{Glu} + [Glu]} \tag{16}$$

$$\frac{d[G6P]}{dt} \cdot \frac{1}{[Enzyme]} = \frac{v}{[Enzyme]} = \frac{k_{cat}^{GLUK} \cdot [Glu]}{K_M^{Glu} + [Glu]} \tag{17}$$

$$[Glu] = \frac{K_M^{Glu} \cdot v/[Enzyme]}{k_{cat}^{GLUK} - v/[Enzyme]} \tag{18}$$

Under conditions in which CelB is inhibited, the steady-state glucose concentration at the active site of GLUK is 3-times higher for co-encapsulation than for separate encapsulation. In both cases glucose concentration is two orders of magnitude below K_m and an increased local substrate concentration should accelerate turnover proportionally. Surprisingly, this would also be the case under conditions in which CelB is not inhibited, at a steady-state glucose concentration 5-fold below K_m. However, although the concentration of the glucose intermediate is also significantly below K_m, the measurements indicate that co-encapsulation of CelB and GLUK does not positively affect overall turnover. This discrepancy could be either due to an artifact in the determination of the kinetic parameters or the observed effect might not be attributable to the model introduced above. Unfortunately, the K_m and k_{cat} values for glucose phosphorylation by GLUK-CelB-P22 were not included in the publication.

The observed rates listed in Table 1 are reflected in the slopes of the progress curves presented in the supporting information of reference [34]. Perplexingly, the graph for uninhibited CelB, which the authors concluded does not show channeling, exhibits characteristics of substrate channeling in case of CelB-GLUK co-encapsulation as judged by the shorter transient time to reach the steady state. Similarly, although less clearly, the transient time seems to be shorter for CelB-GLUK-P22 in

the inhibited case. Regardless of whether channeling occurs or not, the steady-state-rate should be limited by the activity of CelB and thus be equal in both cases. The increased slope is thus more like attributable to an accelerated rate for CelB catalysis in the case of co-encapsulation than to glucose channeling between the two enzyme active sites.

Table 1. Experimental and derived kinetic parameters for the CelB/GLUK-catalyzed cascade reaction in P22.

	no acceleration by co-encapsulation		**acceleration by co-encapsulation (CelB inhibited)**	
	Separated	Co-Encapsulated	Separated	Co-Encapsulated
k_{cat}(GLUK) [s^{-1}]	49	52	52	47*
K_m(GLUK) [mM]	0.68	0.66	0.9	0.84*
v/E_0 observed [s^{-1}]	7.6	7	0.52	0.16
$[Glu]_{steady\ state}$ [mM]**	0.125	0.103	0.009	0.003

*Values were determined for glucose phosphorylation by CelB-GLUK-P22, whereas GLUK-CelB-P22 was used in the experiment to assay the cascade reaction. **$[Glu]_{steady\ state}$ [mM] refers to the apparent glucose concentration at the active site of GLUK, which would be higher than in bulk if channeling occurred.

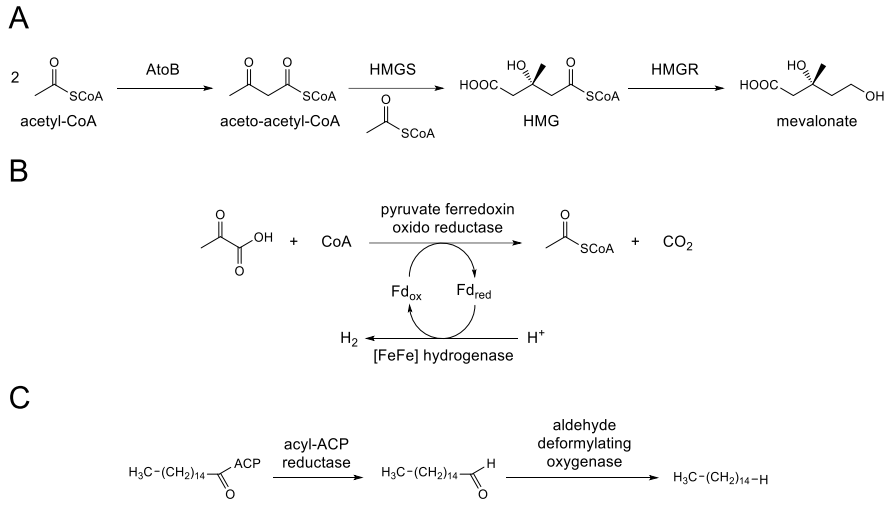
1.6 In vivo enzyme co-localization for metabolic engineering

When it comes to enhancing metabolic flux in microorganisms by engineered enzyme co-localization, the gap between enthusiasm and success is illustrated by comparing the number of reviews [13,17,106–111] to the number of articles describing conceptually novel approaches [16,112–114]. In contrast to in vitro experiments, in vivo studies make no attempt to provide experimental evidence for actual substrate channeling, but content themselves with comparing metabolite titers. In addition to enhancing higher overall reaction rates and skewing the partitioning of branch point intermediates into a desired direction, the maintenance of high concentration gradients that can drive thermodynamically unfavorable reactions, is of particular interest in metabolic engineering [115] – all of which can be achieved by co-localizing enzymes of a pathway.

1.6.1 Protein scaffolds

Dueber et al. designed an artificial synthetic scaffold by genetically fusing repeats of three different protein binding domains from rodents and co-expressed this construct with three consecutive enzymes of the mevalonate pathway (Scheme

16A) that were each C-terminally fused to peptides, which bind to one of the binding domains each [112]. The number of repeats for each domain was varied systematically with the optimal configuration resulting in a 77-fold increase in mevalonate titer (to approximately 0.6 g/L) compared to enzymes without scaffold. To demonstrate that this artificial scaffold is not limited to a single pathway, three enzymes required to produce glucaric acid were similarly co-localized, which increased yields 3-fold and following optimization overall 5-fold [116]. The level of clustering by such a scaffolding approach is likely linked to the quaternary structure of the involved enzymes. If one of them forms oligomers, the linear scaffolds become cross-linked and might form a dense network containing many copies of all relevant enzymes.



Scheme 16. Metabolic pathways improved by protein co-localization. A) Upper mevalonate pathway. Claisen condenstation of two equivalents of acetyl-CoA is catalyzed by acetoacetyl-CoA thiolase (AtoB). Condensation of another equivalent acetyl-CoA yielding 3-hydroxy-3-methylglutaryl-CoA (HMG) is catalyzed by 3-hydroxy-3-methylglutaryl-CoA synthase (HMGS). Reduction of HMG by HMG reductase (HMGR) affords mevalonate, a key intermediate in the biosynthesis of isoprenoid building blocks. B) Electrons produced during pyruvate decarboxylation are transferred to a [FeFe] hydrogenase via the electron transport protein ferredoxin (Fd). C) Formation of pentadecane by coupled reduction of acyl-acyl carrier protein (acyl-ACP) and oxidative deformylation of the resulting aldehyde.

1.6.2 Nucleic acid scaffolds

Delebecque et al. designed RNA molecules to serve as discrete, one-dimensional and two-dimensional scaffolds including two different types of aptamer domains [113]. The discrete scaffold is formed by a single RNA molecule folded into a duplex. The template for the one-dimensional scaffold is constructed from another single

RNA molecule that dimerizes to form a tile, many of which self-assemble into 200-300 nm long RNA nanotubes. Another two different designed RNA-strands assemble into a two dimensional net with ~100 nm diameter. The functionality of co-localization was demonstrated by tethering the two halves of split GFP using corresponding aptamer binding domains on the different scaffolds. For each scaffold, but not in the unscaffolded case, fluorescence was turned on. Co-localizing *Clostridium acetobutylicum* [FeFe]-hydrogenase and *Spinacia olearcea* ferredoxin in such a manner improved overall hydrogen production [117]. Pyruvate ferredoxin oxidoreductase from *Desulfovibrio africanus* was co-expressed, but not tethered (Scheme 16B). In comparison to the unscaffolded enzymes, the discrete, one-dimensional and two-dimensional scaffolds increased the hydrogen production 4-, 11-, and 48-fold respectively. Scaffolds with disabled aptamer sites served as a control. With ferredoxin being a redox-shuttle rather than an enzyme, this system is conceptually similar to the in vitro DNA-scaffold covalently tethering NAD^+/NADH (see section 1.5.3, page 52ff).

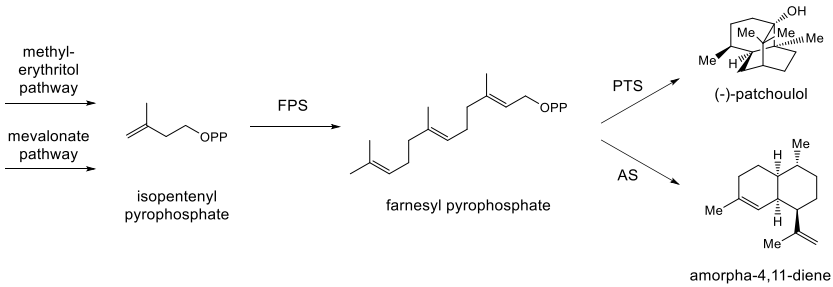
Enablement of diffusive substrate channeling by these RNA scaffolds was showcased by an 2.4-fold increase in pentadecane yield upon co-localizing acyl-acyl carrier protein reductase and aldehyde deformylating oxygenase (Scheme 16C) [118]. In this work the authors demonstrated the importance of enzyme orientation. Changing the angle between the two active sites by introducing an additional base pair in a stem loop reduced the 2.4-fold improvement by 40 % to 1.5-fold. Tethering four enzymes involved in the biosynthesis of malate (a succinate precursor) from pyruvate and carbon dioxide onto a two-dimensional scaffold led to an 88 % increased succinate titer relative to expressing unscaffolded enzymes [118].

Complementary to the RNA constructs, DNA was also employed as scaffold [114]. The same three enzymes of the mevalonate pathway described above were fused to different zinc finger DNA binding domains and co-expressed with a plasmid bearing various DNA sequences that serve as scaffold. DNA-scaffolding resulted in a ~2.5-fold increase in mevalonate titer. Using a similar strategy to co-localize coumaroyl-CoA ligase and stilbene synthase the titer of polyphenol resveratrol was increased five-fold, while scaffolding methylglyoxal synthase, 2,5-diketo-D-gluconic acid reductase and glycerol dehydrogenase increased production of 1,2-propanediol from dihydroxyacetone up to 4.5-fold.

1.6.3 Cell organelles as scaffolds

Fast-growing plants like tobacco are interesting hosts for the production of high-value natural products such as terpenes. To that end, farnesyl pyrophosphate synthase (FPS) and patchoulol synthase (PTS) were overexpressed in tobacco with and

without an N-terminal tag directing the enzymes to chloroplasts, yielding 30 and 5 µg patchoulol per gram fresh weight leaves respectively (Scheme 17) [119].



Scheme 17. Biosynthesis of sesquiterpenes (-)-patchoulol or amorpha-4,11-diene in plants. Depending on the subcellular location, the general terpene precursor isopentenyl pyrophosphate is either synthesized via the methyl-erythritol pathway (plastids) or the mevalonate pathway (cytosol). Cyclization of farnesyl pyrophosphate by terpene cyclases patchoulol synthase (PTS) or amorpha-4,11-diene synthase (AS) afford the sought-after fragrance molecule (-)-patchoulol or anti-malaria drug precursor amorpha-4,11-diene, respectively.

The actual production is likely higher because a considerable share of the volatile product is lost by evaporation. Terpenes originating from the cytosol and plastids can be distinguished by different isotope-labeling patterns upon feeding [1-^{13}C]-glucose. Isopentenyl pyrophosphate (IPP) derived from the cytosolic mevalonate pathway contains up to three labels as opposed to two labels when originating from the plastidic methyl-erythritol pathway. The different labeling patterns were used to determine which cellular compartment the sesquiterpene had originated from. The generalizability of this approach was demonstrated by similarly increasing amorpha-4,11-diene accumulation and a moderate improvement in *R*-limonene yields. Although not discussed by the authors, the higher yields tapping the plastidic over the cytosolic isoprenoid pool might originate from higher local concentration of enzymes and intermediates within the chloroplasts. Alternatively, different regulation of the upstream enzymes filling the IPP pool for the two pathways might be responsible.

Stephanopoulos and co-workers used the natural yeast mitochondrium to compartmentalize the entire metabolic pathway producing short branched-chain alcohols such as isobutanol, isopentanol and 2-methyl-1-butanol [16]. The pathway for isobutanol synthesis from pyruvate is divided into an upstream part, consisting of three enzymes that are naturally localized to the mitochondrium and two downstream enzymes situated in the cytosol. The authors hypothesized that diffusion of the intermediate α-ketoisovalerate, or the corresponding amino acid valine formed

by transamination, across the mitochondrial membrane limits overall pathway efficiency. Indeed, when the downstream enzymes were equipped with an N-terminal mitochondrial localization signal and the entire pathway overexpressed in yeast, isobutanol titers improved by 260 % as opposed to 10 % when the overexpressed downstream enzymes resided in the cytosol.

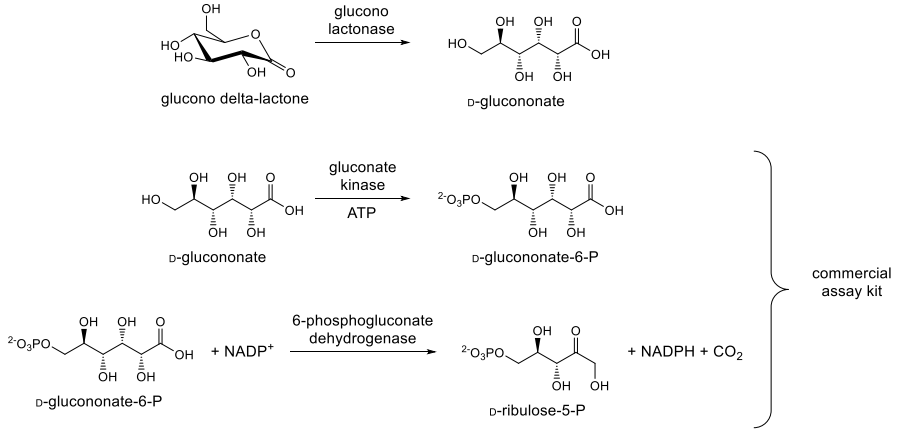
1.7 Conclusions

Over the last decade, many artificial scaffolds have been developed to co-localize enzymes and thus mimic natural multi-enzyme complexes [120]. In contrast to the thorough kinetic investigations and academic debates—often spanning years if not decades—to decide whether substrate channeling actually exits in individual natural systems, the characterization of engineered systems is often relatively rudimentary, relying on a single (and sometimes inappropriate) method to render a verdict as to whether or not substrate channeling is induced by the scaffold. Natural systems rely on intricate constructions such as internal tunnels within proteins or highly evolved electrostatic channels on their surfaces to guide short-lived intermediates between active sites. Complementarily, bacterial microcompartments and (putative) metabolons span between 100 and 1,000 nm [121] and consist of densely packed enzymes. Since nature went to such great lengths for effectuating substrate channeling, it may be necessary to implement one if not both of these design principles to create artificial metabolons.

In order to truly advance our understanding of the conditions that favor channeling, characterization of engineered constructs should rely on at least two methods to analyze the purported effect. Furthermore, careful control experiments need to be conducted to determine the kinetic parameters of the interrogated enzymes in the same setting as they are placed in the alleged channeling situation. Only then can tethering, surface interactions and the like be excluded as the cause of any observed acceleration or deceleration [122,123].

Probably because glucose oxidase and horseradish peroxidase are commercially available in high quantities and fluorogenic substrates are abundant, these enzymes are among the most popular systems to test for channeling induced by artificial scaffolds. In order to eliminate alternative effects as the origin of an observed rate enhancement for the cascade reaction it should become standard operational procedure to assay the impact of co-localization on the catalytic parameters of each individual enzyme. Peroxidase activity is obviously easily assayed by addition of exogenous H_2O_2 and any of a plethora of fluorogenic or chromogenic substrates. Slightly more tricky, but nevertheless feasible by virtue of commercially available kits, glucose oxidase activity can be assayed by quantifying the glucono-*δ*-lactone oxidation product using coupled enzymatic reactions in addition to detecting the

hydrogen peroxide byproduct (Scheme 18). In that case, care has to be taken to ensure that the coupling reactions are not rate limiting. However, ideally, one should choose model cascade reactions for which each individual step can be monitored directly. Lastly, competition experiments with exogenously added catalase, which is also commercially available, pose little additional experimental effort and can corroborate the conclusions drawn from transient kinetic analyses.



Scheme 18. Coupled enzyme assay for quantification of the glucono-δ-lactone product of glucose oxidase. Hydrolysis of the lactone occurs spontaneously at neutral pH [124] or – if necessary – can be accelerated by addition of gluconolactonase [125]. The bioanalytical determination of D-gluconate is frequently used in the food industry, such that the required enzymes and reagents are readily available as kits from commercial suppliers.

The improvements in product yields upon co-localizing enzymes of various pathways in vivo are moderate, yet significant. In general, the findings are in line with a computational study on small molecule substrate channeling performed by Wingreen and co-workers (section 1.3.2.6, page 34) [35]. The only exception is the 77-fold increase in mevalonate production reported by Dueber et al., the origin of which has not been scrutinized and might not be exclusively rooted in shorter diffusive pathways of the two intermediates, but also improved enzyme stoichiometries. One of the most striking effects from in vivo co-localization to date was not obtained by clustering enzymes, but by covalently tethering a redux shuttle in proximity to a redox enzyme [113].

1.8 Aims of this thesis

The overarching goal of this thesis was to explore the potential of the capsid-forming microbial enzyme, lumazine synthase, as well as engineered and evolved

variants thereof, to serve as a host for enzyme cascade reactions. We used a simple encapsulation strategy to construct artificial nanoreactors and investigate whether co-encapsulation facilitates substrate channeling. As a kinetic model system an aldolase and an alcohol dehydrogenase were selected because small molecule channeling can be readily probed with this enzyme pair by a variety of strategies, including introduction of competing catalysts, direct measurement of transient times, and determination of the steady-state concentration of a fluorescent intermediate. Directed evolution experiments conducted in the Hilvert lab over the last few years have spawned a family of artificial aldolases spanning a wide range of kinetic parameters. This variety can facilitate dissection of the factors influencing intermediate channeling. Because practical applications of aldolases are often limited by enzyme inhibition and cytotoxicity of the aldehyde products or intermediates, installing coupled reactions that provide or consume these metabolites in situ could potentially ameliorate this problem in a simple manner.

The central enzyme in this study is the computationally designed and evolved retro-aldolase, RA95.5-8, which catalyzes the aldol cleavage of 4-hydroxy-4-(2-napthyl)butan-2-one with a rate acceleration (k_{cat}/k_{uncat}) of 3×10^8 over background. Chapter 2 describes a detailed mechanistic characterization of this catalyst. The aim of these studies was to identify the rate-limiting step for the aldol and retro-aldol reaction in order to guide efforts for further enzyme optimization as well as its implementation in capsid-enclosed enzymatic cascade reactions. Prior to studying the effects of encapsulation on this enzyme, the mechanism by which protein nanocontainers derived from a lumazine synthase from the hyperthermophilic bacterium *Aquifex aeolicus* encapsulate positively charged cargo proteins was investigated in detail. As described in Chapter 3, increased fluorescence energy transfer efficiency between fluorescent proteins upon encapsulation was exploited to probe the kinetics of host-guest association and the stability of the resulting complexes. Drawing from the insights gathered in this study, Chapter 4 presents strategies for targeting enzymes to the lumen of these protein nanocontainers to assess potential kinetic benefits of co-confining the retro-aldolase described in Chapter 2 with an alcohol dehydrogenase. Finally, as an outgrowth of this work, Chapter 5 examines an alternative approach to confine single proteins within the thermostable capsid shell of wild-type lumazine synthase.

New tools for constructing hierarchical protein assemblies can contribute to the advancement of white biotechnology by making existing biochemical pathways more robust and providing a means to install completely new ones. Recently, considerable interest has been directed at substrate channeling, but disentanglement of various contributing factors remains difficult and attribution of true substrate

channeling challenging. To that end, this thesis not only describes a thorough characterization of one selected encapsulation system, but also introduces new tools to investigate functional host-guest complexes generated by other strategies. These efforts will hopefully reduce the ambiguity in future experimental studies and thus contribute to efforts to determine the principles that govern metabolic channeling. Such insights will facilitate the development of generalizable blueprints for the robust design of functional components for synthetic biology.

2 Mechanistic insights into a computationally designed and evolved aldolase

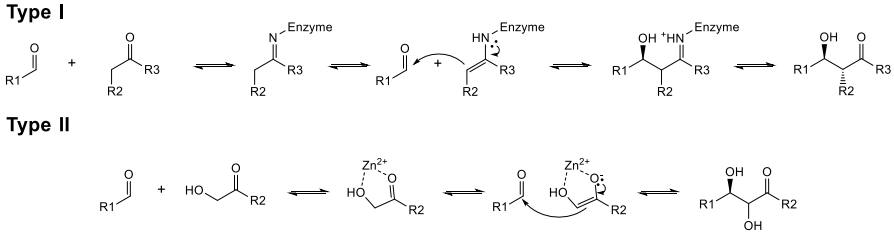
2.1 Introduction

2.1.1 Natural aldolases

Aldolases are among the few biocatalysts that can form carbon-carbon bonds (while installing up to two new stereocenters) and are thus potentially synthetically useful for constructing molecular scaffolds and for enantioselective synthesis of α-substituted secondary alcohols [126]. These enzymes catalyze the nucleophilic attack of an enolizable donor on a carbonyl acceptor (Scheme 19). Natural aldolases are primarily involved in carbohydrate metabolism and considerable effort has been invested toward broadening their substrate scope. Nevertheless, it remains challenging to evolve natural aldolases toward accommodating donors and acceptors that are very distinct from the metabolites involved in carbohydrate conversion [127].

Scheme 19. Aldol reaction catalyzed by aldolase enzymes.

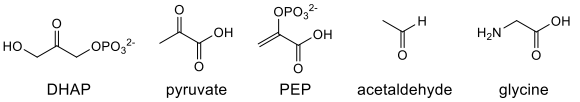
Aldolases are classified into type I and type II enzymes according to their mechanism. Type I aldolases, found primarily in plants and mammals, employ a lysine residue to activate the donor by forming a covalent Schiff base (or imine) intermediate. The latter tautomerizes to a nucleophilic enamine, which is poised for attack on an aldehyde. Subsequent hydrolysis of the resulting iminium group regenerates the catalytic amine (Scheme 20) [128]. Type II aldolases, predominating in bacteria and fungi, stabilize an enolate by coordination to a Zn^{2+} ion [129].



Scheme 20. Reaction mechanisms of type I and type II aldolases.

If the catalyzed carbon-carbon bond fission is thermodynamically more favorable than formation, the responsible aldolases are sometimes referred to as retro-aldolases. However, the terminology is completely arbitrary and does not confer any information about mechanism or substrate scope.

Although aldolases are often promiscuous with respect to the acceptor substrate, they are generally specific for a particular donor. The enzymes are thus subdivided into four main types based on their preferred donor substrate: dihydroxyacetone phosphate (DHAP), pyruvate or phosphoenolpyruvate (PEP), acetaldehyde, and glycine (Scheme 21) [126]. The recent discovery that fructose-6-phosphate aldolase from *E. coli* [130] and mutants thereof [131,132] accept non-phosphorylated dihydroxyacetone, hydroxyacetone, 1-hydroxy-butanone and several 2-hydroxy-aldehydes as donors considerably expands the palette of starting materials that are compatible with natural aldolases.



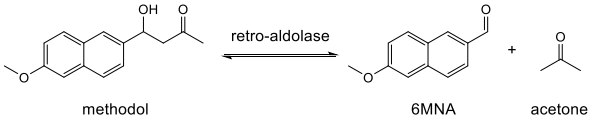
Scheme 21. Most common donors accepted by natural aldolases.

A notable example for relaxation of donor substrate specificity is the redesign of the phosphate binding site in L-rhamnulose-1-phosphate aldolase [133]. Mutating an asparagine to an aspartate residue increased the initial reaction rate for the aldol reaction between dihydroxyacetone and Cbz-protected (*S*)-2-aminopropanal more than two-fold.

2.1.2 A computationally designed aldolase

To demonstrate the prowess of computational enzyme design [11] and to widen the scope of aldol reactions that can be catalyzed enzymatically, David Baker and co-workers designed many retro-aldolases for the cleavage of (*S*)-methodol to 6-methoxy-naphthaldehyde (6MNA) and acetone (Scheme 22) [134]. The structures of the rate-limiting transition states for four different catalytic motifs were optimized using quantum mechanical calculations. The resulting theozymes comprise the relevant substrate functional groups as well as potential catalytic residues [135]. Since the retro-aldol reaction is a multi-step reaction containing multiple transition states, composite transition states were designed and placed in silico into the binding pockets of 71 different proteins scaffolds. Variation of the orientation of the theozyme as well as placement and conformations of the participating catalytic side chains within a given pocket were scored with respect

to catalytic geometry and steric clashes, followed by optimization of the identity and conformation of surrounding, non-catalytic side-chains. The genes of 72 designs were synthesized and expressed, of which 32 were catalytically active. The best variants catalyzed the retro-aldol reaction with $k_{cat} \approx 10^{-4}$ s^{-1} and a catalytic rate enhancement, $k_{cat}/k_{uncat} \approx 2 \times 10^4$.



Scheme 22. Reaction catalyzed by computationally designed retro-aldolases.

An improved design procedure with different weights for the individual reaction steps contributing to the composite transition state yielded catalysts with similar catalytic rate enhancements, yet a higher overall success rate for active computational designs (75 % versus 44 %) [136]. The weights for the transition states for carbinolamine formation and breakdown of the tetrahedral intermediate accompanied by liberation of a water molecule were increased relative to that for carbon-carbon bond cleavage. Among these designs was RA95, created by introducing 11 mutations into a $(\beta\alpha)_8$ TIM barrel scaffold. Cassette mutagenesis of active site residues improved k_{cat}/K_m of this variant about 16-fold (RA95.4). A slightly improved variant, RA95.5, contains a total of six mutations relative to the original computational design (RA95.0), which confer a 73-fold improvement in overall activity. This improvement is attributable in part to the T83K mutation, which introduces a new functional group into the active site [137,138]. Interestingly, the accumulated mutations also inverted the enantioselectivity of the catalyst, favoring cleavage of (*R*)-methodol by 3:1. Eight additional rounds of directed evolution using error-prone PCR and DNA shuffling to create molecular diversity subsequently yielded variant RA95.5-8, which exhibits a 60-fold higher k_{cat}/K_m due to 8 additional mutations (Figure 20A). Analysis of the crystal structures of variants RA95.0, RA95.5 and RA95.5-5 covalently modified with the di-ketone analog of methodol gave an unexpected surprise. In contrast to the original design, in which the designed catalytic lysine, Lys 210, forms a Schiff base adduct with the inhibitor, RA95.5 was modified twice, once at lysine 210 and additionally at lysine 83. The most evolved variant was exclusively modified at lysine 83 (Figure 20B).

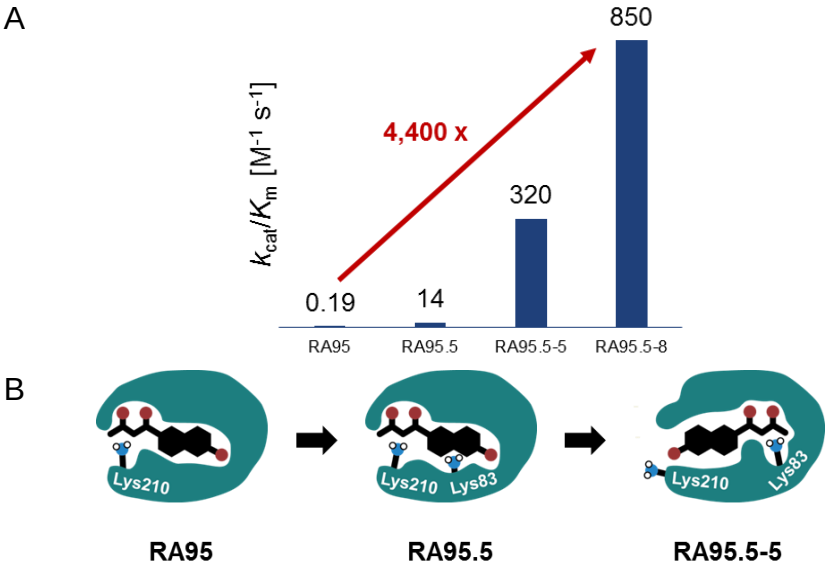


Figure 20. Directed evolution of retro-aldolase RA95. A) Catalytic proficiency for different RA95 variants. B) Introduction and re-purposing of lysine 83 in different RA variants.

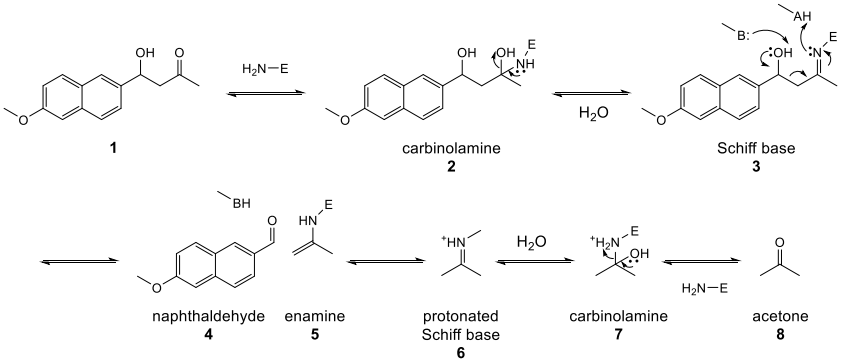
While the first 5 rounds of directed evolution starting from RA95.5 yielded a 23-fold improved catalytic proficiency, the last 3 rounds from RA95.5-5 to RA95.5-8 only afforded a two-fold increase in k_{cat}/K_m. Presumably, additional improvements require sampling of a larger sequence space, which exceeds the capacity of the medium-throughput essay employed. In order to identify the step in the catalytic mechanism, which limits the overall reaction, we carried out detailed kinetic studies of the most evolved variant, RA95.5-8, with the objective of guiding future evolution efforts. Furthermore, since natural aldolases predominantly process molecules without distinct spectroscopic properties, the fluorogenic and chromogenic properties of methodol and 6-methoxy-2-naphthaldehyde studied herein might shed light on the mechanisms of aldolases in general.

2.2 Results

2.2.1 Identification of potential rate-limiting steps

The combination of multiple elemental steps with transition states stabilized by different environments poses a particular challenge for the design of efficient retro-aldolases (Scheme 23). Following binding of the β-hydroxy ketone substrate **1** in a hydrophobic pocket, the substrate carbonyl group is set up for nucleophilic attack by a deprotonated amine—the reactive lysine(s) in RA95. Besides favorable alignment of the substrate and amine, including an unobstructed Bürgi-Dunitz trajectory [139], the rate of this step is determined by both the pK_a of the lysine, and thus the fraction of unionized amine, as well as the polarization of the carbonyl group, which can be increased by coordination to either a Lewis or Brønsted acid. In the transition state leading to the formation of carbinolamines like **2**, the developing negative charge on the carbonyl oxygen is generally stabilized by one or more hydrogen bond donors in a so-called oxyanion hole [140]. The tetrahedral carbinolamine can collapse by either expelling the amine and regenerating substrate **1**, or by eliminating water and forming an iminium ion. Deprotonation of the latter generates Schiff base **3**. Which of the two reaction steps leading to imine **3** is rate limiting depends on the pH [141]. Deprotonation of the catalytic amine is a prerequisite for nucleophilic attack and thus favored at high pH. Elimination of hydroxide, on the contrary, is limited by protonation by a general acid and thus favored at low pH. In enzymes, this predicament is usually resolved by lowering the pK_a of the catalytic amine to physiological pH, e.g. by positioning it next to a positively charged residue or immersing it in a hydrophobic environment. Formation of Schiff base **3** is thus a likely candidate for the rate-limiting step.

Another candidate is the actual chemical step, cleavage of the Cα-Cβ bond of intermediate **3**. Protonation of the imine group creates an electron sink that can take up the electrons in the breaking bond. The reaction is further driven by partial or complete deprotonation of the hydroxyl group in the transition state. Concerted action by proton acceptors and donors, plus appropriate alignment of the relevant orbitals constitute a complex set of requirements that may render this step overall rate limiting. While dissociation of naphthaldehyde **4** and tautomerization of enamine **5** to the corresponding Schiff base should be rapid, hydrolysis of the Schiff base might be relatively slow. The electrophilicity of the imine is substantially enhanced upon protonation. However, the same environment that leads to reduction of the catalytic amine's pK_a might analogously depress that of the iminium ion and thus slow down formation of carbinolamine **7** to the extent that product release becomes rate-limiting.



Scheme 23. Detailed reaction mechanism for retro-aldol cleavage of methodol.

Knowing which of the three potential candidates is the actual rate-limiting step will guide further improvement of the RA95.5-8 retro-aldolase. If formation of Schiff base **3** were the slowest step, substrate binding as well as positioning of its carbonyl group with respect to the reactive amine will need to be optimized. If instead Cα-Cβ bond fission is limiting, efforts should focus on coordination of the β-hydroxy group and the Schiff base nitrogen, together with stabilizing a favorable geometry of this intermediate. Finally, if product release is the actual bottleneck in catalysis, measures to increase the pK_a of the iminium ion will be necessary.

2.2.2 Product release

The dissociation constant of the Schiff base formed between acetone and the reactive amine, $K_{D,\ acetone}$, was determined to be 420 ± 90 mM by measuring the ability of acetone to inhibit retro-aldol cleavage of (±)-methodol under competitive inhibition conditions (Figure 21). Acetone concentrations exceeding 400 mM resulted in decreased k_{cat} values indicating that inhibition was no longer solely competitive.

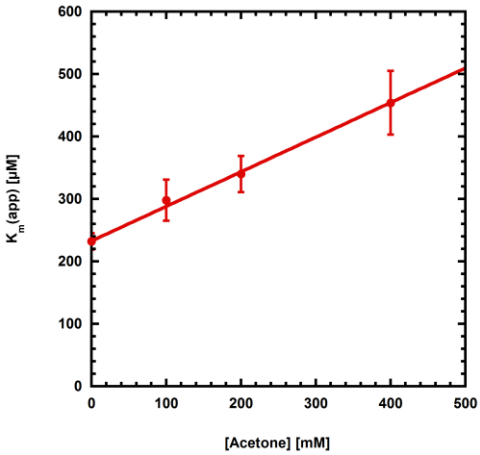
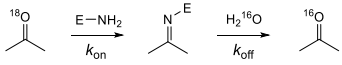


Figure 21. Determination of the dissociation constant of the Schiff base between acetone and retro-aldolase. The apparent K_m for (±)-methodol was determined at different concentrations of acetone by fitting the initial rates of 6MNA formation as a function of (±)-methodol concentration to the Michaelis-Menten equation: $v_o/[E]_o = k_{cat}[S]/(K_m + [S])$. Linear regression of the apparent K_m values against [acetone] using $K_m(app) = K_m + [acetone]/K_I$ yields the inhibition constant K_I, which is equivalent to the dissociation constant of the inhibitor.

The rates of imine formation and hydrolysis between acetone and lysine 83 were determined by isotope dilution of the carbonyl oxygen (Scheme 24). 90 % ^{18}O-labeled acetone was synthesized by incubating [^{16}O]-acetone in [^{18}O]water. The ratio of [^{18}O]-acetone to [^{16}O]-acetone was determined by their relative ion counts using GC-MS.



Scheme 24. Oxygen-isotope exchange by amine catalysis.

The background reaction in aqueous solution proceeds via hydration of the carbonyl group and is catalyzed by H^+, OH^- and any buffer component. Since the rate constants for phosphate catalyzed hydrations have been reported [142], the experiment was conducted in phosphate buffer (25 mM NaP_i, 100 mM NaCl, pH 7.5). The rate constant for background isotope exchange was determined to be 0.310 ± 0.006 h^{-1}, which is in acceptable agreement with the rate of 0.192 h^{-1} estimated from rate constants reported in reference [142] (Figure 22A).

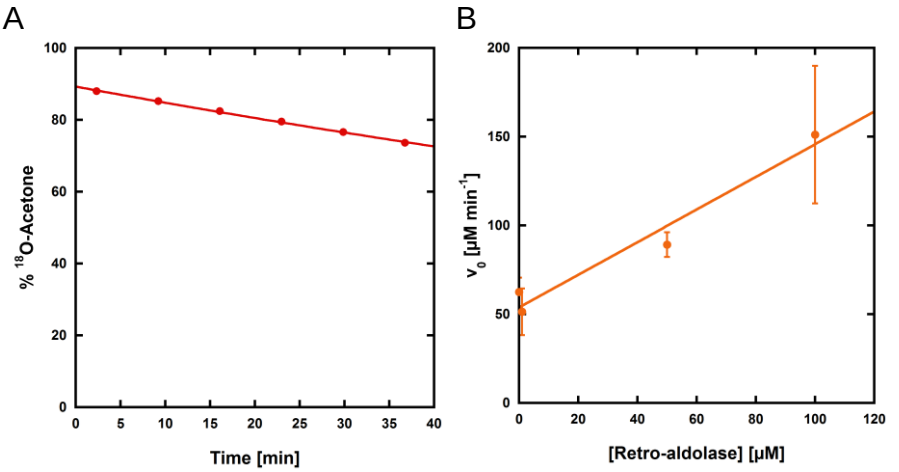


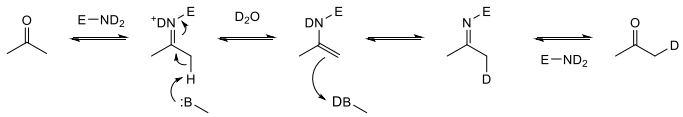
Figure 22. Oxygen isotope exchange in acetone. A) Time dependent ^{18}O-depletion in [^{18}O]-acetone upon incubation in 25 mM phosphate buffer. B) Rate of isotope exchange in [^{18}O]-acetone as a function of enzyme concentration. From the slope (0.92 ± 0.13 min^{-1}) and the concentration of acetone in the experiment (10 mM) the second-order rate constant k_{cat}/K_m was estimated to be 92 ± 13 M^{-1} min^{-1}.

In order to determine the second-order rate constant for retro-aldolase-catalyzed isotope exchange, [^{18}O]-acetone was incubated with various concentrations of enzyme and the rate of ^{18}O-isotope dilution was determined (Figure 22B). Owing to a K_D in the high millimolar range, the enzyme could not be saturated with acetone, precluding direct determination of k_{cat}. However, knowing K_D and the substrate concentration, one can use equations (1) and (2) to determine the ratio of on- and off-rate constants for covalent modification of the enzyme with acetone.

With $K_{D,\,acetone}$ ≈ 420 mM and acetone concentration, [S] = 10 mM, the majority of the enzyme is in the free state since $K_D/[S] = [E]/[ES] \approx 42$ (re-arrangement of the binding equation, eq. (19)). Since $[E]/[ES] = k_{off}/(k_{on} \bullet [S])$, we know that the first-order rate constant k_{off} is approximately 42-times higher than the pseudo-first order constant $\{k_{on} \bullet [S]\}$ for acetone binding and thus Schiff base formation must be rate-limiting. Consequently, the measured overall rate of exchange, 0.92 ± 0.13 min^{-1}, must correspond to $\{k_{on} \bullet [S]\}$, which at [S] = 10 mM yields k_{on} = 92 ± 13 M^{-1} min^{-1}. From k_{on} and K_D, k_{off} is determined by $K_D \bullet k_{on}$ to be 39 ± 10 min^{-1}.

$$K_D = \frac{k_{off}}{k_{on}} = \frac{[E][S]}{[ES]} \leftrightarrow \frac{K_D}{[S]} = \frac{k_{off}}{k_{on}[S]} = \frac{[E]}{[ES]} \qquad (19)$$

The rate of enamine formation can be determined by hydrogen isotope exchange with acetone. This reaction was monitored by ^{1}H-NMR and ^{2}H-NMR for incorporation of deuterium into h_6-acetone in D_2O or protium into d_6-acetone in H_2O, respectively (Figure 23) [143].



Scheme 25. Amine-catalyzed hydrogen isotope exchange in acetone.

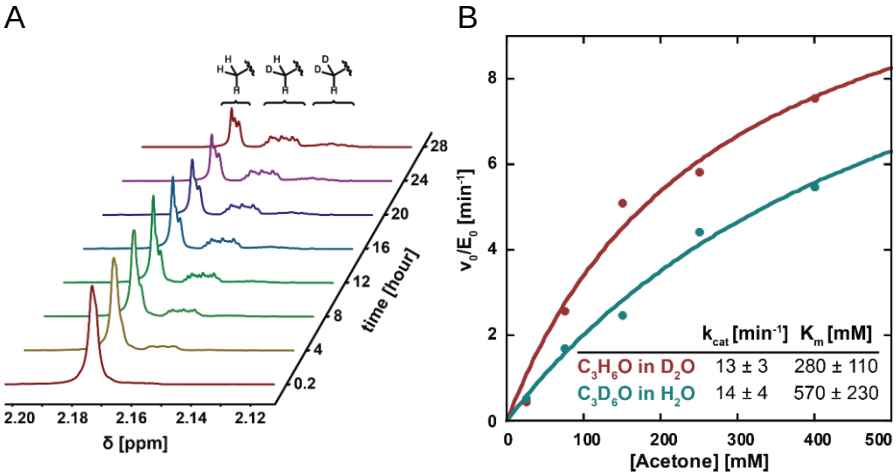


Figure 23. Hydrogen isotope exchange in acetone. A) 600 MHz ^{1}H NMR spectra of a mixture of 100 µM RA95.5-8 and 400 mM h_6-acetone in D_2O recorded at different time points. The three multiplets with decreasing chemical shifts correspond to the signal from un-, mono- and di-deuterated methyl groups. The multiplicity of each peak increases with time, reflecting 4-bond coupling with deuterium incorporated into the other α-methyl group in symmetric acetone. B) Michaelis-Menten plot for hydrogen isotope exchange of h_6-acetone in D_2O (red) and d_6-acetone in H_2O (teal) catalyzed by RA95.5-8.

Given that imine formation and hydrolysis proceed faster than α-hydrogen exchange, the k_{cat} values of ~14 min^{-1} determined for retro-aldolase-catalyzed hydrogen exchange in acetone either correspond to deprotonation of the imine α-carbon or protonation of the enamine. The multi-step mechanism and the solvent isotope effect confound the analysis. Given that the enamine-imine equilibrium between primary amines and aliphatic ketones usually lies on the side of the imine, enamine formation is likely rate-limiting [144] and therefore a kinetic isotope effect for hydrogen abstraction should be observed. However, this effect might be obscured by

altered pK_a values [145] and protein dynamics [146] in heavy water. The solvent isotope effect on k_{cat} for methodol cleavage k_H/k_D has been determined to be 2.60 ± 0.22 for RA95.5-8. Consequently, the exchange of protium at the methyl group of acetone in ^1H_2O is certainly higher than in the two cases described above and enamine-imine tautomerization cannot be rate limiting for the retro-aldol reaction (k_{cat} ≈ 10 min^{-1}). Hence, we can conclude that product release does not limit the RA95.5-8-catalyzed retro-aldol cleavage of methodol.

2.2.3 6-Methoxy-2-napthaldehyde binding

Because the strong fluorescence of methodol and 6-methoxy-2-napthaldehyde obscure any potential changes in RA95.5-8 tryptophan fluorescence, binding of 6-methoxy-2-napthaldehyde was followed by measuring the decrease of 6MNA fluorescence using a stopped-flow apparatus.

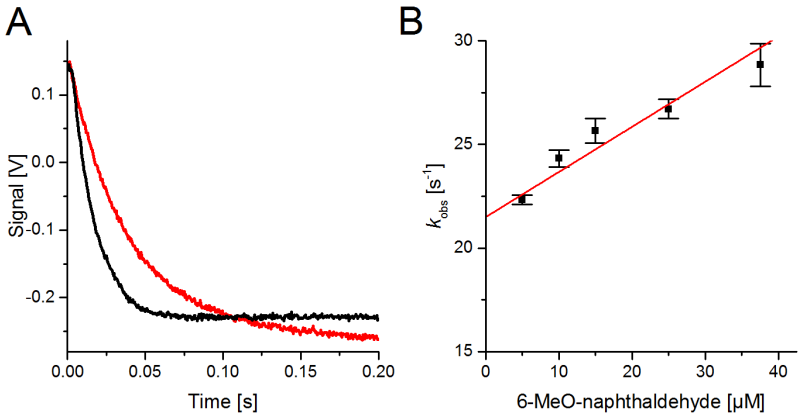


Figure 24. 6-Methoxy-2-naphthaldehyde binding to RA95.5-8. A) Decrease in fluorescence upon mixing 6-methoxy-2-naphthaldehyde (final 10 µM) with enzyme (final 5 µM) in the presence (red) and absence (black) of 500 mM acetone (red). The linear phase following the exponential phase reflects aldehyde consumption by methodol formation. B) Observed rate constant for substrate binding in the presence of 500 mM acetone as a function of substrate concentration.

The rate constant obtained for binding is about 3-fold lower in the presence of 500 mM acetone as opposed to no acetone, because the 4 % acetone co-solvent decreases the polarity of the solvent and thus reduces the hydrophobic interactions driving substrate binding. In the presence of 500 mM acetone, the parameters for 6MNA binding to RA95.5-8 were determined to be k_{on} = 0.22 ± 0.03 µM^{-1} s^{-1}, k_{off} = 21.5± 0.4 s^{-1} and K_D = 98 ± 13 µM (Figure 24).

2.2.4 Active site mutations

In order to estimate the importance of the original catalytic amine, lysine 210, on the nucleophilicity of lysine 83, the catalytic amine in RA95.5-8, several active site mutants were prepared. The K210R mutation reduced k_{cat} approximately two-fold, whereas replacing this lysine with a methionine increased catalytic proficiency about two-fold, primarily by reducing K_m (Table 2 and Figure 25).

Table 2. Michaelis-Menten parameters for RA95.5-8 mutants. The error corresponds to the error of the fit.

	k_{cat} [min^{-1}]	K_m [µM]	k_{cat}/K_m [M^{-1} min^{-1}]
RA95.5-8	10.5 ± 0.9	230 ± 50	45,000 ± 10,000
RA95.5-8 K210M	11.9 ± 0.8	119 ± 23	100,000 ± 20,000
RA95.5-8 K210R	6.2 ± 0.3	340 ± 30	18,000 ± 2,000

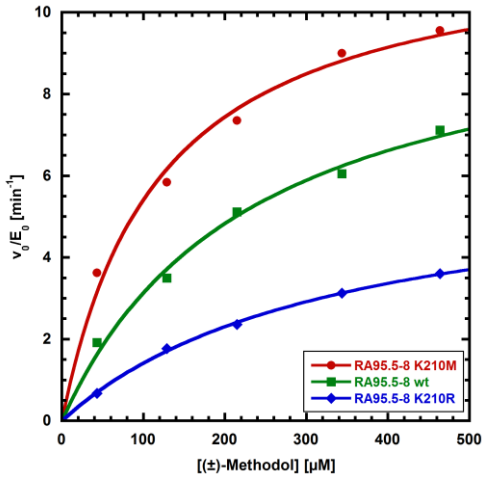


Figure 25. Michaelis-Menten plots for RA95.5-8 active site mutants.

The roles of the two lysines were further investigated. Replacing lysine 83 with a methionine residue decreases the enzyme's proficiency by three orders of magnitude, clearly indicating that this amino acid has taken over the role of reactive amine from lysine 210 (Figure 26A). On the other hand, replacing lysine 210 with a methionine surprisingly reduces the inflection point of the pH-rate profile by almost one unit (Figure 26B). Since the inflection point most likely corresponds to the

pK_a of the catalytic amine, this result suggests that the presence of a second lysine at the active site does not depress the pK_a of the catalytic lysine via Coulombic interactions, but rather increases it. Alternatively, the similarity of the pH-rate profiles of variants RA95.5-8 and RA95.5-8 K83M might indicate that catalysis is dependent on the protonation state of lysine 210 even though this is not the catalytic residue. Potentially, protonation of lysine 210 could switch the enzyme from an active to an inactive state or cause a conformational change that prevents substrate from binding. The idea of such a conformational equilibrium is supported by the K210M mutation primarily affecting K_m and thus substrate binding and/or the events preceding the rate-limiting step. Sufficiently high substrate concentrations would still push the equilibrium toward the reactive state. In this hypothesis the proton on the spectator lysine acts as a competitive inhibitor (Figure 27).

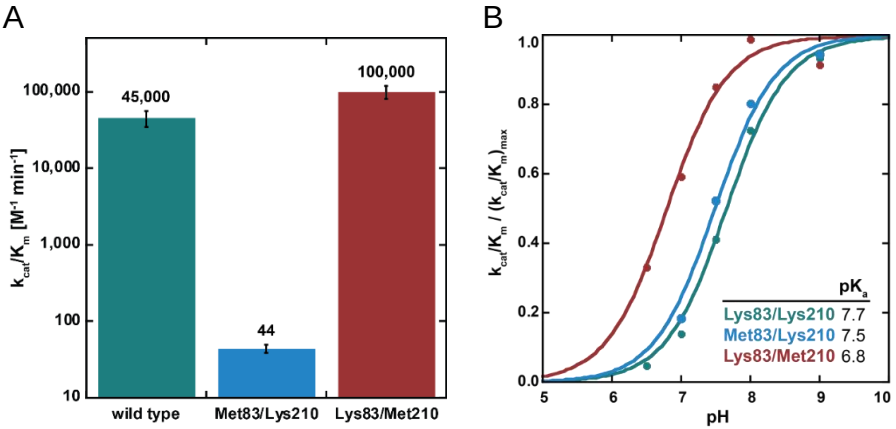


Figure 26. Probing the role of active site lysine residues. A) Catalytic proficiency of different aldolase variants at pH 7.5. The three order of magnitude reduced k_{cat}/K_m value for Met83/Lys210 indicates that Lys83 is the reactive amine. B) pH-rate profile for methodol cleavage. The pK_a value of the reactive Lys83 (red) is almost one unit lower than the inflection point for the wild type enzyme (green).

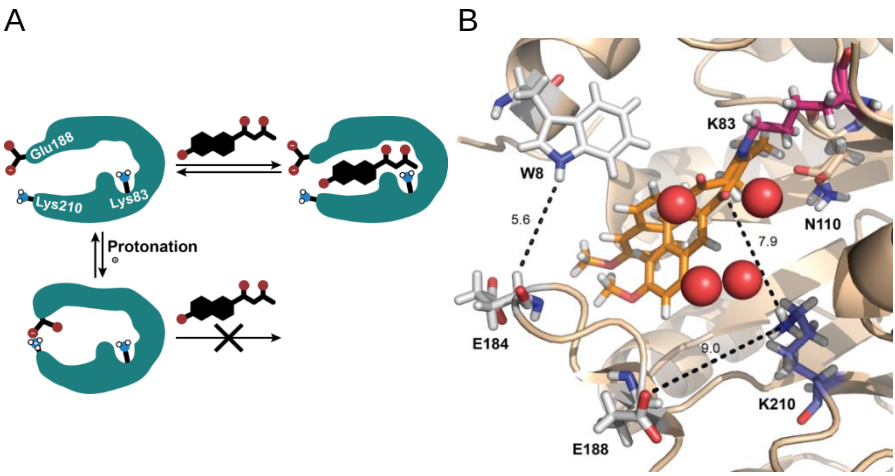
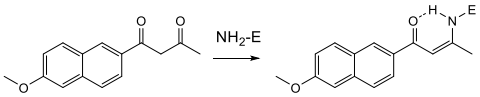


Figure 27. Putative conformational pre-equilibrium governed by protonation of lysine 210. A) Illustration of how protonation of lysine 210 might affect substrate binding to RA95.5-8. B) Active site of RA95.5-5. Proximity of lysine 210 to the loop spanned by residues I179-I189 could influence its conformation upon amine protonation, e.g. by modulating the interaction with one of the two glutamate residues. Figure was created using PyMOL [54] and PDB structure 4A2R [137].

Attempts to crystallize RA95.5-8 and RA95.5-8 K210M with and without a covalent adduct with diketone **9** (Scheme 26) were unsuccessful. Crystallization trays for these variants were set up in the same fashion as RA95.5-8 T53L/K210H and related mutants. However, only RA95.5-8 T53L/K210H formed diffraction quality crystals [147]. The reluctance to crystallize might be rooted in the flexibility of the loops bracketing the active site. In the structure of RA95.5-8 T53L/K210H, the I179-I189 loop was disordered and could only be partially resolved.

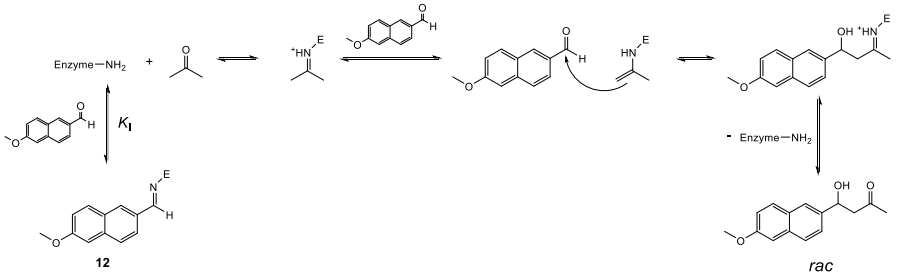


Scheme 26. The reactive lysine of retro-aldolases can be covalently modified with a 2,4-diketone derivative of methodol.

2.2.5 Aldol reaction

Taking advantage of an on-fluorescence screening assay, retro-aldolase activity in RA95.5-8 was evolved by monitoring the cleavage of methodol. For practical applications, however, running the aldol reaction in the synthetic direction is of greater interest. Unfortunately, the computationally designed aldolases examined

to date have been generally poor catalysts for aldol synthesis, suffering from substrate inhibition by 6-methoxy-2-naphthaldehyde. Since the selection for high activity also resulted in a catalyst with substantial total catalytic turnovers [137], this inhibition became presumably suppressed. We therefore examined the ability of RA95.5-8 to catalyze the stereoselective aldol reaction between 6-methoxy-2-naphthaldehyde and acetone (Scheme 27).



Scheme 27. Mechanism of the synthetic aldol reaction and competing substrate inhibition.

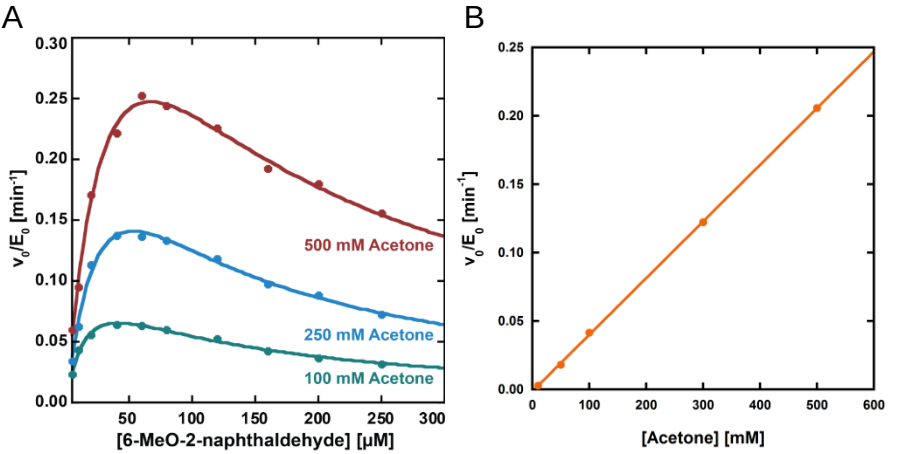
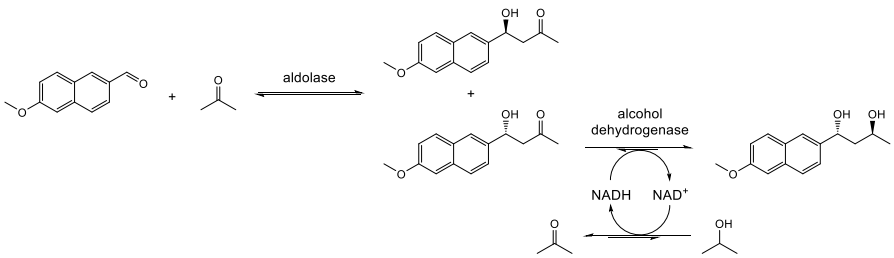


Figure 28. Synthetic aldol reaction catalyzed by RA95.5-8. A) Initial reaction rate of methodol formation in dependence of 6-methoxy-2-naphthaldehyde concentration at different acetone concentrations. B) Initial reaction rate in dependence of acetone concentration in presence of 160 µM 6-methoxy-2-naphthaldehyde.

As shown in Figure 28, RA95.5-8 does promote methodol synthesis from 6-methoxy-2-naphthaldehyde and acetone. Nevertheless, the initial rate plots clearly indicate that substrate inhibition by the aldehyde remains a problem. The inhibitory constants for 6-methoxy-2-naphthaldehyde increase with increasing acetone

concentration, indicating competitive inhibition (Figure 28). Although the imine formed between the aldehyde and the reactive lysine is particularly stable because it is in conjugation with the aryl group, its formation is still reversible. Attempts to reductively trap the imine formed between methodol or acetone and the reactive lysine (**3**) with $NaBH_3CN$ or $NaBH_4$ were unsuccessful. Instead, only the mass corresponding to reduced imine **3** was found in ESI mass spectra, both for RA95.5-8 and RA95.5-8 K210M. These results reflect the 5,000 lower dissociation constant of the imine formed between lysine 83 and aldehyde as compared to acetone. Failure to trap either the imine formed with methodol or acetone indicates that the reactions following their formation, carbon-carbon cleavage and hydrolysis, respectively, proceed faster than reductive trapping using the relatively weak reducing reagents compatible with aqueous conditions.

The methodol formed by RA95.5-8 was isolated after the reaction reached equilibrium (circa 40 % conversion with respect to 6-methoxy-naphthaldehyde in the presence of 500 mM acetone). Although RA95.5-8 exhibits moderate stereoselectivity for the cleavage of methodol (a 16:1 preference for (*R*)- over (*S*)-methodol), chiral analysis showed that the purified product was racemic. This unexpected result can presumably be explained by the reversibility of the thermodynamically unfavorable synthetic reaction. When acetone was replaced with 2-butanone as the aldol donor, the rate of reaction was considerably diminished and an unequal distribution of diastereomers was obtained. The observed selectivity in this case may reflect either an inherent stereochemical preference by the enzyme or the fact that the reaction had not been run to equilibrium. The diastereomers have not been assigned yet.



Scheme 28. Aldolase/alcohol dehydrogenase cascade reaction to pull the equilibrium to towards 1,3-difunctionalized products.

The enantiomeric excess might be improved if the synthetic reactions were stopped at earlier times. Alternatively, the equilibrium of the reaction could be shifted by coupling the synthetic aldol reaction to a second step. One possibility—

elimination of water to form the α,β-unsaturated ketone—is already inherently catalyzed by RA95.5-8. In another approach, an alcohol dehydrogenase specific towards methyl ketones might stereoselectively reduce the carbonyl group, installing a second stereocenter, and thus pull the overall reaction toward products. As described for other biocatalytic applications of dehydrogenases, isopropanol could be used to regenerate the reduced co-factor and produce additional acetone donor (Scheme 28).

2.3 Conclusion and outlook

The evolutionary trajectory that produced RA95.5-8 is remarkable and fuels reveries that tailor-made biocatalysts with useful rates and selectivities may soon be at hand. Although computational design can to date provide artificial enzymes with only low activities, these catalysts represent valuable starting points for subsequent improvement by directed evolution [148–152]. The greatly enhanced activity of RA95.5-8 is testimony to the power of this approach. Nonetheless, this artificial enzyme is still far away from the proficiency and selectivity of natural enzymes. Considerable inhibition of RA95.5-8 by 6-methoxy-2-napthaldehyde highlights a crucial feature of natural aldolases, namely their strict specificity for a particular donor. Aldehydes are generally very reactive toward nucleophilic attack, a desirable feature in their role as aldol acceptors, but a great problem if they compete with the aldol donor for the catalytic amine. In the case of RA95.5-8, the K_i for acetone is about 4,000-fold higher than that for 6-methoxy-2-naphthaldehyde, such that the productive Schiff base adduct with acetone can only be significantly populated in the presence of a vast excess of donor. While the lack of selectivity observed for the synthetic reaction is problematic, this promiscuity also presents an opportunity. The flexibility of RA95.5-8's activity site and the availability of two reactive lysine residues are promising features that might be exploited for evolution of catalysts with new substrate specificities or altered reaction pathways. The reality of this vision has already been demonstrated by the ability of a RA95.5-8 double mutant to stereoselectively catalyze conjugated additions [147].

With respect to methodol, substrate inhibition and moderate enantioselectivity can be improved by redesigning the activity site into a pocket that snugly fits around the methodol substrate and restricts the orientation of bound 6-methoxy-2-napthaldehyde such that its carbonyl cannot be attacked by the reactive lysine. Obexer et al. undertook this challenge, employing high-throughput droplet-based microfluidics to screen millions of diversified clones. The best catalyst, variant RA95.5-8F, is about 60-fold more active than RA95.5-8 and exhibits considerably reduced inhibition by 6-methoxy-2-napthaldehyde, enabling efficient, stereoselective synthesis of (*R*)-methodol [153]. Characterization of this variant using the

methods described in this chapter and comparison with RA95.5-8 will provide a solid basis for understanding the structural changes responsible for the achieved improvements. These efforts will be complemented by more detailed substrate binding studies and single-turnover kinetic experiments using stopped-flow spectroscopy. A narrow band pass filter should allow selective detection of subtle changes in protein fluorescence against the background of intense substrate and product fluorescence.

2.4 Materials and methods

2.4.1 Molecular biology

Using plasmid pET29b_RA95.5-8 [137] as template, mutations K83M, K210M, and K210R as well as the double substitutions K83M/K210R and K83M/K210M were introduced by overlap extension PCR using flanking primers RA95.5_fw and RA95.5_rv together with primers containing the corresponding mutations (Table 3). The resulting gene products were digested with *Nde*I and *Xho*I and ligated into correspondingly digested and dephosphorylated pET29b_RA95.5-8.

Table 3. Oligonucleotides used for molecular biology.

Name	Sequence
K210M_f	CGTGGTGAAAGTCGCGATGCTGGACATTAGTGAGCG
K210MR_r	CGCGACTTTCACCACG
K210R_f	CGTGGTGAAAGTCGCGCGTCTGGACATTAGTGAGCG
K83M_fw	TTACGCGGTTGGCCTGTCTATTATGACAGAAGAGAAATACTTTA ATGGCAGCTATG
K83M_rv	TAATAGACAGGCCAACCGCGTAA
RA95.5_fw	TGATAACATATGCCGCGTTATCTGAAAGG
RA95.5_rv	TTATCACTCGAGTGATCCCTCGATTAATTCTTT

2.4.2 Retro-aldolase steady state kinetic measurements

Acetonitrile (MeCN) and (±)-methodol dissolved in acetonitrile were added sequentially to 25 mM HEPES, 100 mM NaCl, pH 7.5 and equilibrated at 29 °C using a Peltier system for temperature control. The final concentration of MeCN was 2.7 %. The reaction was initiated by addition of 0.2 µM enzyme in a final volume of 600 µL. Product formation was monitored by absorption at 350 nm ($\Delta\varepsilon$ = 5,970 M^{-1} cm^{-1}). Methodol substrate concentrations were determined by absorption at 330 nm using an extinction coefficient of 1,420 M^{-1} cm^{-1}. The pH-rate profiles were determined in 50 mM bis-Tris propane, 100 mM NaCl over the range pH 6.5-9.0.

2.4.3 Isotope exchange experiments

Preparation of 25 mM HEPES, 100 mM NaCl, pD 7.5 in D_2O. 477 mg HEPES and 100 mM NaCl were dissolved in 5 mL D_2O and lyophilized to exchange protium to deuterium. The lyophilized powder was dissolved in 60 mL D_2O and pH adjusted to pH 7.1 using NaOD. A correction of 0.41 pH units is necessary to convert the operational pH determined with a standard glass electrode to pD [154]. This buffer was

used to determine the solvent isotope effect for retro-aldolase catalyzed methodol cleavage and acetone hydrogen exchange.

Hydrogen exchange in acetone. Purified RA95.5-8 in H_2O buffer was exchanged 1,250-fold with D_2O buffer using ultrafiltration spin tubes (10 kDa MWCO). The reaction mixture contained 100 µM RA95.5-8, h_6-acetone at concentrations between 25 and 400 mM and 200 mM h_3-MeCN in a final volume of 600 µL. The reaction was started by addition of enzyme and monitored on a 600 MHz NMR spectrometer operated by Dr. Marc-Olivier Ebert. The concentration of acetone α-protons was determined by integrating their signal and comparison with the signal for MeCN used as an internal standard. The initial rates for protium loss were fitted to the Michaelis-Menten equation. For the inverse experiment, d_6-acetone and h_3-acetone were dissolved in H_2O buffer (supplemented with 2.5 % D_2O to lock the field frequency) and the loss of deuterium in acetone was monitored by ^2H NMR.

[^{18}O] exchange in acetone. 90 % ^{18}O-acetone was prepared by mixing 22 µL acetone (0.3 mmol) with 70 µL $H_2{}^{18}O$ (3.86 mmol, 97 % ^{18}O) and incubating for four days at room temperature, yielding 3.3 M acetone with 90 % ^{18}O. The background reaction was monitored by incubating 100 mM [^{18}O]-acetone in 25 mM NaP_i, 100 mM NaCl, pH 7.5 and injecting aliquots on a GC-MS. The ratio of [^{18}O]-acetone to [^{16}O]-acetone was determined by comparing the ion-count at 60 and 58 m/z. To monitor the enzyme-catalyzed reaction (10 mM [^{18}O]-acetone in 25 mM NaP_i, 100 mM NaCl, pH 7.5 and 1-100 µM RA95.5-8), aliquots were removed at various time points and filtered using ultrafiltration spin tubes (10 kDa MWCO) and the filtrate immediately injected into the GC-MS.

2.4.4 Naphthaldehyde binding

A solution containing 10 µM RA95.5-8 and 500 mM acetone in 25 mM HEPES, 100 mM NaCl, pH 7.5 was mixed with solutions containing 500 mM acetone and varying concentrations of 6-methoxy-2-naphthaldehyde in 25 mM HEPES, 100 mM NaCl, pH 7.5 in a 1:1 ratio in a SX18-MV stopped-flow spectrometer (Applied Photophysics, UK) equilibrated with a water bath at 25 °C. The decrease in fluorescence upon excitation at 330 nm was followed using a long pass filter with an edge at 400 nm (Applied Photophysics, UK). Progress curves were fitted to equation (20). The k_{obs} values were fitted as a function of final 6-methoxy-2-naphthaldehyde concentration to equation (21) to obtain k_{on} and k_{off}.

$$y = a \cdot e^{-k_{obs}t} + b \cdot t + c \tag{20}$$

$$k_{obs} = k_{off} + k_{on}[S] \tag{21}$$

2.4.5 Aldol reaction

Acetonitrile (MeCN), acetone and 6-methoxy-2-naphthaldehyde dissolved in acetonitrile were added sequentially to 25 mM HEPES, 100 mM NaCl, pH 7.5 and equilibrated at 29 °C using a Peltier system for temperature control. The final concentration of MeCN was 2.7 %. The reaction was initiated by addition of 1.0 µM enzyme in a final volume of 600 µL. Product formation was monitored by absorption at 350 nm ($\Delta\varepsilon$ = -5,970 $M^{-1}\,cm^{-1}$). The initial slopes were fitted to equation (22) as a function of 6-methoxy-2-naphthaldehyde concentration.

$$\frac{v_0}{[E]_0} = \frac{k_{cat}[S]}{K_m + [S]\left(1 + \frac{[S]}{K_I}\right)} \tag{22}$$

3 Properties of an artificial protein container

3.1 Introduction

3.1.1 Construction principles of viral capsids

Symmetry is a concept that can be understood and appreciated by the broad population irrespective of culture and profession – by a ballet dancer as much as an aerospace engineer, by an architect as much as a baker. Besides being aesthetically pleasing (or boring), man-made symmetry is often employed with a very pragmatic goal: to conserve effort. It requires less planning and fewer different parts to construct a symmetric building or vehicle. Nature employs symmetry for the same reason. Less information is needed to build up two symmetric organs than two different ones. Apparently, the loss in functional diversity (imagine you had one eye for day and another eye for night vision) is outweighed by simplicity and the smaller amount of information that must be stored. Maintenance and duplication of nucleic acids is resource intensive, so reducing genome size can provide a competitive advantage.

Optimizing information storage efficiency is taken to the extreme in viruses, tiny parasites that replicate the small number of genes they possess by exploiting the resources of organisms they infect. Viruses protect their sensitive genetic information by encapsulating it in sturdy protein shells. These shells, which are often referred to as capsids [155,156], are hollow spheres composed of multiple, periodically arranged copies of the same protein. This arrangement allows viruses to encode their entire shell by a single gene. In their seminal work *Physical Principles in the Construction of Regular Viruses* [157], which is to physical virologists what *The Origin of Species* [158] is to evolutionary biologists, Donald Caspar and Aaron Klug describe a theory for the assembly of helical and spherical viruses that is still generally accepted today. This theory was formulated following the discovery that all spherical viruses, regardless of origin, have cubic symmetry, more specifically icosahedral symmetry, and are composed of identical subunits with molecular weight around 20 kDa. Within the three types of symmetry in which all directions in 3-dimensional space are equivalent – tetrahedral, octahedral and icosahedral – icosahedral symmetry can arrange the highest number of identical subunits equivalently on a sphere (60 versus 12 and 24 for tetrahedral and octahedral symmetry, respectively). In other words, a shell of a particular size can be constructed from the smallest subunits (requiring the shortest genes) if icosahedral symmetry is employed. Sixty subunits arranged in this fashion form a dodecahedron (composed of 12 pentamers containing 5 subunits each) or an icosahedron, assembled from 20 trimers. These two Platonic bodies are the inverse of each other, with the faces of one occupying the positions of the vertices in the other. Whether one regards the

trimer or the pentamer as the capsomer, i.e. the geometric building block that can be assembled into a regular three-dimensional body and usually comprises multiple elementary subunits, is geometrically arbitrary and often based on which low-order assembly the subunits adopt outside the capsid (Figure 29).

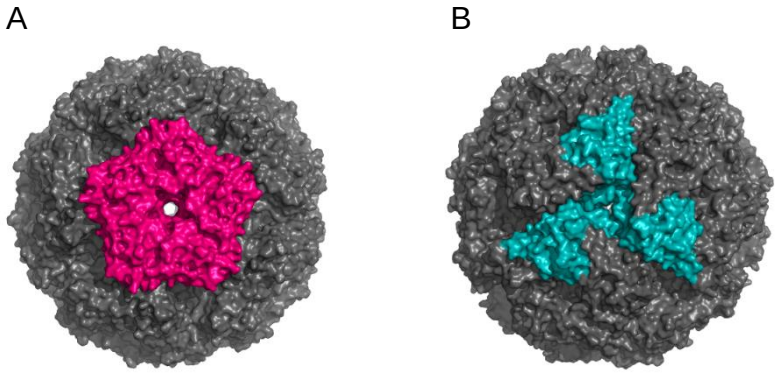


Figure 29. Crystal structure of the 60-meric lumazine synthase from *Aquifex aeolicus* (PDB 1HQK [159]) with a pentameric (A) and a trimeric (B) capsomer highlighted in color. The regular pentagonal shape of the pentamer, in contrast to the anfractuous topology of the trimer, predestines the former to be the physically relevant capsomer unit.

The conundrum that confronted Caspar and Klug was the observation that viruses are often larger than would be possible if they were only composed of sixty subunits of 20 kDa each. The problem is the impossibility to arrange more than 60 subunits in a shell such that all subunits are identical. The solution proposed by Caspar and Klug evoked a new paradigm, which was quickly adopted by the physical virology community: quasi-equivalence. In this framework, the interactions between subunits are not strictly identical, but almost equivalent, if slightly deformed, when interactions between subunits situated around different symmetry axes are compared. In Caspar and Klug's words, *molecular structures are not built to conform to exact mathematical concepts but, rather, to satisfy the condition that the system be in a minimum energy configuration* [157]. Drawing from Linus Pauling's work on antigen-antibody interactions [160], they postulated that the non-equivalency required to realize icosahedral bodies with more than 60 subunits would be within the margin for the juxtaposition of atoms described by Pauling and declared the interactions quasi-equivalent.

A protein monomer by itself cannot be symmetric. However three asymmetric proteins can be arranged in an equilateral triangle, and these triangles can assemble as planar 2-dimensional lattices in such a way that all units are strictly equivalent, i.e. have exactly the same environment (Figure 30A). Even if the subunits

themselves are not symmetric, the overall lattice is and each unit monomer has exactly identical inter-subunit interactions.

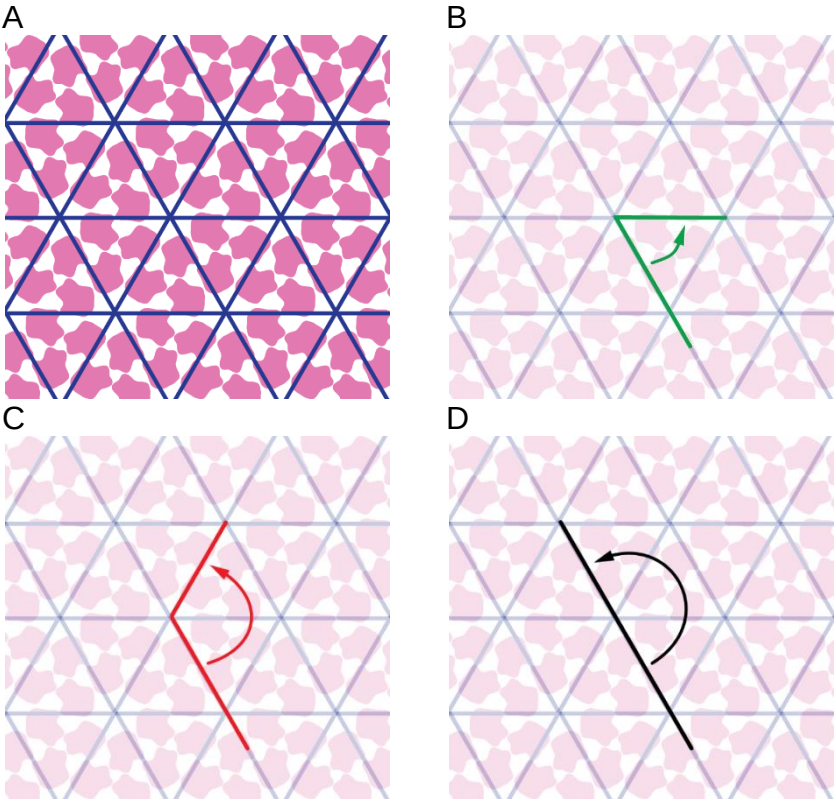


Figure 30. Quasi-equivalence in the construction of viral capsids. A) Two-dimensional hexagonal lattice with asymmetric subunits (pink) in equivalent environments. The lattice has 2-fold, 3-fold and 6-fold rotational symmetry. Cutting the plane and sliding one surface above the other in a 60° (B), 120° (C) or 180° (D) rotation creates a convex surface. Panel A adapted from reference [157].

In order to form the convex surface necessary to obtain a spherical object, one can imagine cutting the plane, and sliding one surface over the other so that the subunits align again exactly (Figure 30B-D). Depending on the angle at which the surfaces are rotated relative to one another, objects with different geometries are obtained. Rotating 60° yields a five-fold symmetric vertex and overall icosahedral symmetry, 120° affords a four-fold symmetric vertex and overall octagonal symmetry, whereas 180° gives a three-fold symmetric vertex and tetrahedral sym-

metry. Obviously, a 60° rotation entails the least deviation from the planar arrangement and best fulfills the criterion of quasi-equivalence. If this geometric operation is performed once on each neighboring hexagon, after 12 steps a closed body comprised of 12 convex pentamers is formed, i.e. a dodecahedron. In this object all subunits are in an identical environment and strictly equivalent. However, if the folding operation is not performed on each hexagon, but some are left out, a larger body can be formed that consists of 12 pentamers (folded) and a variable number of hexamers (unfolded and thus planar). The geometry of the resulting body can be described by the distance between pentamers, represented by the length of vectors h and k, the sum of which connects the centers of the hexamers that were converted into pentameric vertices (Figure 31).

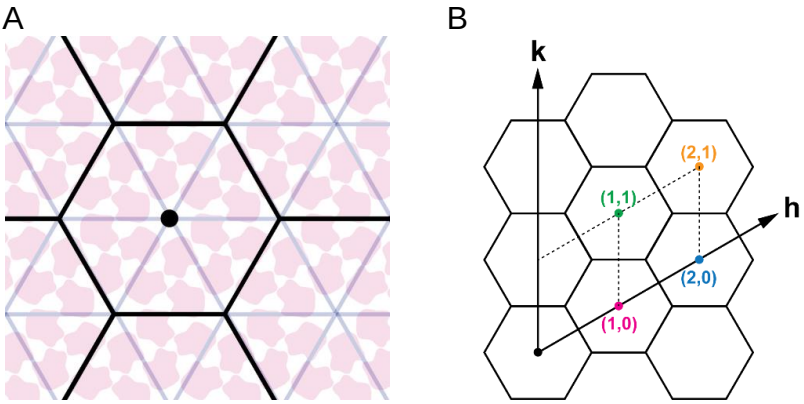


Figure 31. Relation between T number and capsid size. A) The planar trigonal lattice from Figure 30 can be seamlessly divided into regular hexagons. B) The distance between hexagons converted into convex pentameric vertices to enable formation of icosahedrally symmetric capsids can be described by the length of vectors k and h connecting two vertices. Panel B adapted from reference [161].

The resulting icosahedron is described by a T number, where $T = (h^2+hk+k^2)$ and h and k are any integer such that $h \geq 1$ and $k \geq 0$. Each icosahedrally symmetric object described by this formalism consists of 12 pentamers and (T-1) x 10 hexamers. If all the pentameric vertices are neighbors, (h, k) = (1,0) and T=1 and the smallest possible capsid—a dodecahedron—is formed (Figure 32A). For (1,1), or T=3, the next largest object is obtained. It is constructed from 12 pentamers and 20 hexamers. The T=3 truncated icosahedron is also the structure of a C_{60} fullerene or a classic soccer ball. The capsid formed from pentamers related by the (2,0) vector (T=4) is geometrically equivalent to T=1 with each triangle in Figure 31A composed of four smaller triangles. Capsids larger than T=4 are skewed, meaning they are chiral, unless their T number is a multiple of 1 or 3 and a square number (i.e., capsids

with T=9, 16, 25, 27, 75 are not chiral). A T=75 capsid has been proposed for the shell of the carboxysome [162]. Notable examples of natural viruses that deviate from the Caspar and Klug formalism are papilloma and polyoma viruses. Their capsids have overall T=7 geometry, except that the sites usually taken by hexagons are occupied by pentameric capsomers (T=7d) [163]. Consequently, these capsids are composed of 60 subunits fewer than a "normal" T=7 capsid, i.e. (7 x 60) - 60 = 360 monomers.

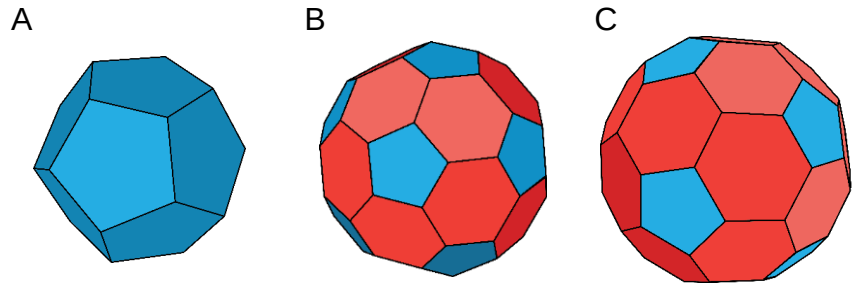


Figure 32. Icosahedrally symmetric capsids with different triangular numbers. A) A T=1 capsid, (h,k)=(1,0), composed of 12 monomers can be represented as a dodecahedron. B) A T=3 capsid, (h,k)=(1,1), composed of 12 pentamers and 20 hexamers gives a truncated icosahedron. C) A T=4 capsid, (h,k)=(2,0), composed of 12 pentamers and 30 hexamers, has the structure of a truncated rhombic triacontahedron.

3.1.2 Capsid assembly and stability

Assembly of viral capsids, often from hundreds of subunits, can to some extent be compared to crystallization, initiated by a nucleation step that is followed by growth [164]. The protein-protein interactions between capsid subunits are dominated by hydrophobic interactions with contact areas that are smaller than generally found in homo-dimeric protein complexes [165]. As a consequence, the association of two subunits is reversible, as is the addition of a monomer to oligomeric capsid fragments. Stable assemblies are only obtained if each component is involved in multiple interactions [166]. Detailed theories about the thermodynamics and kinetics of virus capsid assembly have been developed and tested in vitro as well as in silico, for example by the groups of Adam Zlotnick [167,168] and Michael Hagan [169], who have written excellent reviews on the topic. Although viral capsids have evolved to wrap around their genome, empty shells—so-called virus-like particles—can likewise assemble in vitro. Electrostatics play a crucial role, because the interior surface of viral shells is usually positively charged to be able to interact strongly with the negatively charged nucleic acids packed in the lumenal space. In the absence of DNA or RNA, this positive charge causes the subunits to repel one

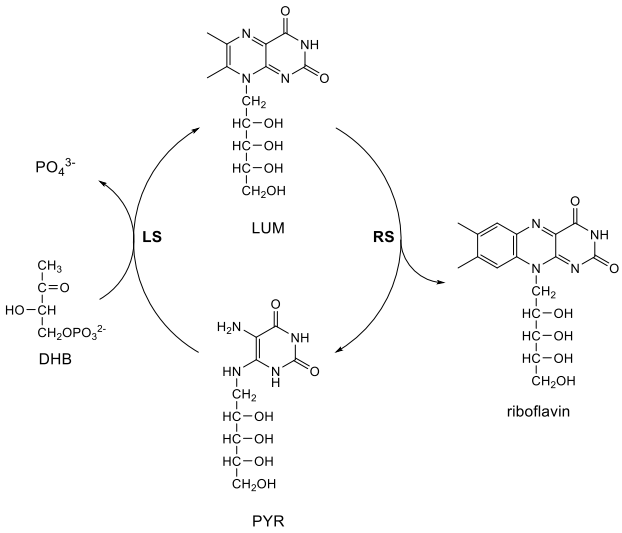
another. As a consequence, in vitro assembly of empty shells often requires high ionic strength to screen the Coulombic interactions and decrease the repulsion [170].

Weak inter-subunit interactions are key to viral capsid assembly. At low concentration, subunits reversibly associate with each other. Above a certain concentration, the critical subunit concentration, association is favored over dissociation, leading to formation of larger assemblies and depletion of free subunits. As the number of binding partners around a subunit increases, the kinetic barrier for its detachment rises, slowing down dissociation and thus driving growth. In this advanced phase of capsid assembly, the solutions become depleted of capsid monomers, which decreases the stability of smaller capsid protein oligomers. Weak inter-subunit interactions allow non-productive intermediates to disintegrate, replenishing the pool of available monomers and avoiding kinetic traps. After completion of the capsid shell, however, detachment of single subunits is associated with a high kinetic barrier. The multiplicity of protein-protein interactions between one subunit and its many neighbors factors exponentially into its dissociation constant. For a weak interaction, with, say, a K_D value of ≈100 μM between triangular capsomers, the dissociation constant for a capsomer from the assembled capsid would be K_D^3 ≈ 1 pM. The associated kinetic barrier renders capsids metastable, i.e. they prevail even if the concentration of their free constituents drops below the critical concentration [167,171]. Similarly, capsids often remain stable under conditions (e.g., ionic strength and pH) that would prevent their assembly. These properties allow viruses to remain intact after lysing their host cell and being extensively diluted in the blood stream or the entire ecosphere. The ability to form a stable spherical object at infinite dilution with a potentially customizable cargo hold makes viral capsids interesting candidates for the design of artificial delivery vehicles and nanoreactors.

3.1.3 The capsid-forming enzyme lumazine synthase

Proteinaceous polyhedra are not only found in viruses, but also constitute the outer shell of bacterial organelles, such as the carboxysome (see Chapter 1) [172]. Smaller analogs include encapsulins [173], which are believed to protect bacteria from oxidative stress, and the lumazine synthase/riboflavin synthase complex [174]. Lumazine synthase (LS) catalyzes the penultimate step in the biosynthesis of riboflavin, namely condensation of 5-amino-6-ribitylamino-2,4(1*H*,3*H*)-pyrimidinedione (PYR) and L-3,4-dihydroxy-2-butanone 4-phosphate (DHB) to give 6,7-dimethyl-8-ribityllumazine (LUM) with concomitant release of inorganic phosphate and water (Scheme 29). Disproportion of two LUM molecules to afford riboflavin and regenerate PYR is catalyzed by the enzyme riboflavin synthase. In fungi and archaea

the two enzymes are found as independent homo-oligomers (lumazine synthase usually exists as a pentamer and riboflavin synthase as a trimer), but in bacteria and plants twelve lumazine synthase pentamers assemble into a dodecahedral 60-meric capsid (T=1), enclosing one riboflavin trimer. The reason for this distinction is unknown. As it impossible to prove that something does *not* exist, similar complexes could conceivably form in fungi and archaea but their assembly might be reversible, potentially orchestrated by the cytoskeleton, and thus not observed upon isolation.



Scheme 29. Reactions catalyzed by lumazine synthase (LS) and riboflavin synthase (RS).

Substrate channeling in the *Bacillus subtilis* lumazine synthase/riboflavin synthase $\alpha_3\beta_{60}$ complex has been investigated by Kis and Bacher [175]. The observation of a lag phase for riboflavin formation for a 1:1 mixture of α_3 and β_{60} but its absence in the case of the $\alpha_3\beta_{60}$ complex suggests substrate channeling in the latter. At high substrate concentrations, the channel becomes leaky, indicating that riboflavin synthase activity cannot keep up with lumazine synthase and LUM escapes into bulk solution. After one of the lumazine synthase substrates is depleted, the rate of riboflavin synthesis drops 30 fold despite high LUM concentrations, reflecting a high apparent K_m value for exogenous LUM that yet has to cross the lumazine synthase shell to be processed by encapsulated riboflavin synthase. The evolutionary

advantage of lumazine channeling is still unclear [174]. At low PYR and DHB concentrations, the overall rate of riboflavin synthesis is greatly enhanced by the $\alpha_3\beta_{60}$ complex and bulk LUM concentrations are kept low. The observation that intermediates in riboflavin biosynthesis, including 6,7-dimethyl-8-ribityllumazine, but not riboflavin bind to the major histocompatibility complex related protein 1 and thus effect T-cell activation [176] suggests a possible explanation why LUM sequestration could be advantageous to pathogens [174].

3.1.4 Engineered lumazine synthase capsids

Lumazine synthase from the thermophilic bacterium *Aquifex aeolicus* (AaLS or AaLS-wt) has been engineered to encapsulate positively charged cargo molecules. Four glutamates were introduced on the lumenal face of each monomer to give the variant AaLS-neg, which encapsulates GFP molecules possessing a deca-arginine tag [177,178]. The capsid's ability to retain proteins tagged with a positively charged peptide was subsequently improved by directed evolution [179,180]. The resulting variant, AaLS-13, has a total of 11 mutations and forms capsids with an increased diameter (35.4 ± 3.2 nm compared to 28.6 ± 2.6 nm for AaLS-neg and 15.4 nm for wild-type AaLS) [159,179]. In contrast to AaLS-wt and AaLS-neg, which are both obtained as assembled capsids upon purification with metal-affinity chromatography, AaLS-13 is purified as a mixture of assembled capsids and smaller fragments (mostly pentamers) [180]. AaLS-13 capsids can be non-covalently loaded with positively charged cargo proteins in vivo and in vitro [179,181].

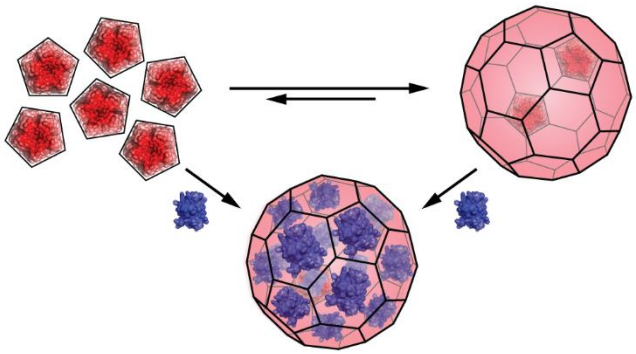


Figure 33. The AaLS-13 encapsulation system. Rational design and directed evolution endowed the lumenal face of the AaLS capsid with substantial negative charge (red). Capsid fragments (largely pentamers) spontaneously assemble to capsids in vitro in the presence of positively supercharged GFP36+ (blue) affording cage complexes. Analogous complexes are obtained upon mixing GFP36+ with assembled AaLS-13 capsids [181]. Since the structure of assembled AaLS-13 is unknown, a hypothetical truncated icosahedron is used to depict the capsid. Figure adapted from reference [181].

For example, mixing positively supercharged green fluorescent protein (GFP36+) [182,183] with capsid fragments affords assembled capsids filled with cargo. Identical guest-capsid complexes are obtained upon mixing guests with pre-assembled capsids (Figure 33). AaLS-13 capsids are even able to incorporate particles as large as ferritin (d = 12 nm) [184], suggesting either an atypical uptake mechanism or partial reversibility of capsid assembly. While the encapsulation of cargo concurrent with capsid formation from fragments is intuitive, the entry of large cargo proteins into assembled capsids is surprising and to our knowledge without precedent.

The microscopic and macroscopic stabilities of capsid-cargo complexes are decisive parameters for biotechnological applications. The development of a positively supercharged protein pair is described in this chapter that enables spectroscopic investigation of the extent of encapsulation by AaLS-13 via changes in Förster resonance energy transfer efficiency (FRET), a phenomenon that has been exploited to characterize other engineered guest-capsid systems [185,186]. This technique is used to gain unprecedented insight into the kinetics of the encapsulation process and the stability of the capsid-cargo complexes.

3.2 Probing guest encapsulation with a protein FRET pair

3.2.1 A protein FRET pair

Introducing four mutations—T65G, S72A, K79R and T203Y (numbering with respect to wild-type GFP)—into GFP36+ [183] afforded TOP36+, named after Topaz, a yellow fluorescent protein (YFP) obtained by introducing the same mutations into wild-type GFP [187]. The absorption maximum of TOP36+ is red-shifted to 513 nm (ε = 69,800 M^{-1} cm^{-1}) versus 488 nm (ε = 36,600 M^{-1} cm^{-1}) for GFP36+. Analogously, whereas GFP36+ displays maximum emission at 508 nm, the emission maximum of TOP36+ is red-shifted to 525 nm. The quantum yield is similar for both variants (0.52 ± 0.03 for GFP36+ and 0.55 ± 0.03 for TOP36+) and independent of the excitation wavelength between 410-450 nm.

Encapsulation of GFP36+ by AaLS-13 capsids [181] has only a marginal effect (2 %) on its fluorescence. In contrast, the fluorescence of TOP36+ is increased up to 10 % upon internalization (Figure 34). It is well known that chloride ions decrease the fluorescence of YFPs [188]. The enhanced fluorescence upon encapsulation is thus likely attributable to a lowered affinity for chloride present in the standard assay buffer (50 mM NaP_i, 200 mM NaCl, 5 mM EDTA, pH 8.0, I = 350 mM), owing to the more negative electric potential within the AaLS-13 capsid.

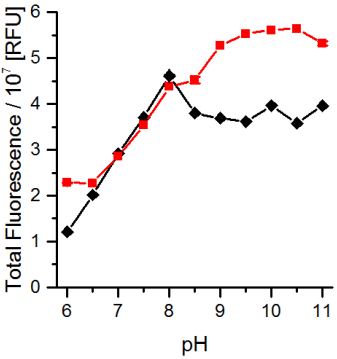


Figure 34. pH-dependence of TOP36+ fluorescence free in solution (black diamonds) and contained in AaLS-13 capsids (red squares).

GFP and YFP have the longest Förster distance and hence the highest FRET efficiency of characterized fluorescent protein pairs [189]. Overlaying the emission and excitation spectra of GFP36+ and TOP36+ reveals that this property is preserved after supercharging (Figure 35A). The emission peak of GFP36+ overlaps with the excitation peak of TOP36+, and while TOP36+ shows little absorption at 400 nm, GFP36+ is still considerably excited. Addition of purified AaLS-13 capsids to such a

mixture increases the emission maximum from 508 to 525 nm upon selective excitation of GFP36+, indicating that co-confinement of the two fluorescent proteins in the AaLS-13 capsid decreases their average distance close to the Förster distance of around 5 Å [189] and thus drastically enhances FRET efficiency (Figure 35B). The most pronounced change is observed for a GFP36+:TOP36+ ratio of 1:4, in accord with the finding that energy transfer efficiency increases upon increasing the number of acceptors per donor [190].

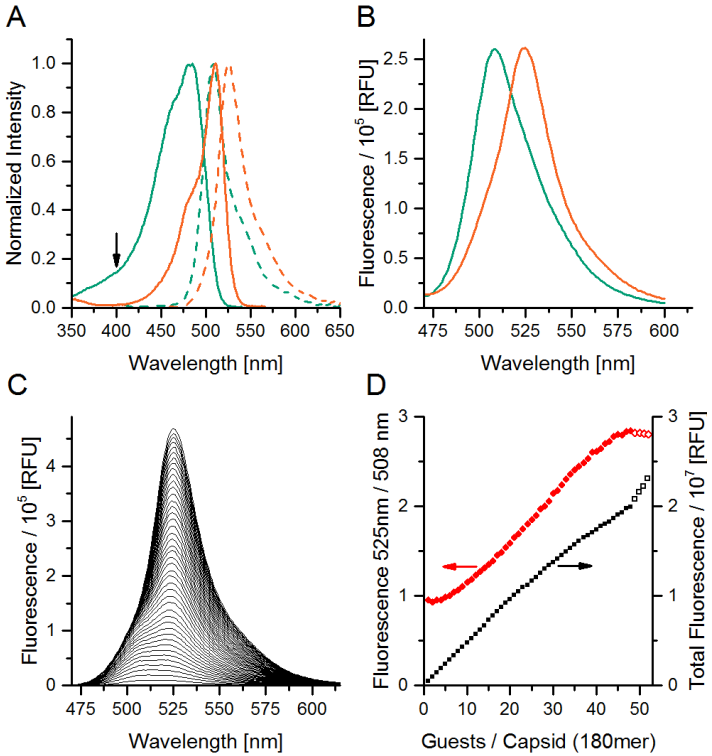


Figure 35. Fluorescence spectra of the GFP36+/TOP36+ FRET pair. (A) Normalized excitation (solid lines) and emission (dashed lines) spectra of GFP36+ (green) and TOP36+ (orange). The arrow indicates the wavelength at which GFP36+ can be selectively excited. (B) Emission spectrum of a 1:4 GFP36+:TOP36+ mixture before (green) and after (orange) encapsulation in AaLS-13 capsids. (C) Emission spectra upon titrating 1 to 42 equivalents of guest into an AaLS-13 capsid (180-mer). After addition of each equivalent a fluorescence spectrum was recorded. (D) Ratio of emission at 525 nm and 508 nm (red diamonds) and total fluorescence (black squares) as a function of guest loading before (filled symbols) and after (open symbols) onset of aggregation.

Although the precise assembly state of the AaLS-13 capsid has not been determined, the capsid diameter observed by negative-staining transmission electron microscopy (TEM) is consistent with symmetric polyhedral capsids having triangulation number T=3 or 4, corresponding to 180 or 240 subunits, respectively [179]. In the absence of detailed structural information and for the sake of simplicity, we report capsid loading assuming a 180-meric capsid structure [181]. Titrating AaLS-13 capsids with a 1:4 mixture of GFP36+ and TOP36+ demonstrates that the extent of FRET depends on capsid occupancy (Figure 35C). Between 10 and 46 guests per 180-mer, the relative emission at 525 nm/508 nm is linearly dependent on the number of guests per capsid and thus a suitable indicator for the extent of guest encapsulation (Figure 35D, red diamonds). At very low loading, donors and acceptors are far apart, whereas at high loading additional effects such as self-quenching become more prominent, resulting in decreased total fluorescence emission per guest (Figure 35D, black squares). Exceeding the loading capacity (45 ± 2 guests per T=3 capsid) leads to aggregation as judged by a sharp increase in light scattering. A decline in overall FRET efficiency and increased total fluorescence per guest (Figure 35D, open symbols) suggest that the additional guests are not encapsulated.

Comparing FRET efficiencies between guests encapsulated in AaLS-13 with values estimated for common capsid geometries could potentially provide insights into the capsid's quaternary state. The ratio of guest to AaLS-13 subunits is easily determined spectrophotometrically. However, since the number of capsid subunits is unknown, so is the number of guests per capsid. While accurate determination of intermolecular distances from FRET between two fluorophores with well separated absorption and emission peaks is fairly straightforward [191], such an analysis is semiquantitative at best in the case of multiple encapsulated protein fluorophores.

For an AaLS-13 capsid filled with a 1:4 mixture of GFP36+/TOP36+ we applied a model proposed by Fábián et al. [190], assuming FRET between one GFP36+ donor and four TOP36+ acceptors (Figure 36A). The average distance between guests is derived from the experimentally observed inner diameter of an AaLS-13 capsid [181], and the solution of the Tammes problem, an approach applied to capsid loading by Cornelissen et al. [186]. Assuming identical capsid diameters and guest:capsid subunit ratios, capsids composed of a higher number of subunits would be more densely loaded and thus exhibit higher FRET (Figure 36B). Although the agreement is imperfect, experimentally observed FRET efficiencies are most compatible with the AaLS-13 capsid being a 240-mer (T=4). This estimate will have to be confirmed by cryo-electron microscopy and native mass spectrometry. Both techniques are now routinely employed to determine capsid composition and structure [192,193].

The deviation at low capsid loading is peculiar and might be explained by a non-statistical distribution of guests among capsids, as previously seen in TEM pictures of AaLS-13 host-guest complexes [181,184], and also observed for solute encapsulation by liposomes [194]. In neither case is the origin of this phenomenon understood [195].

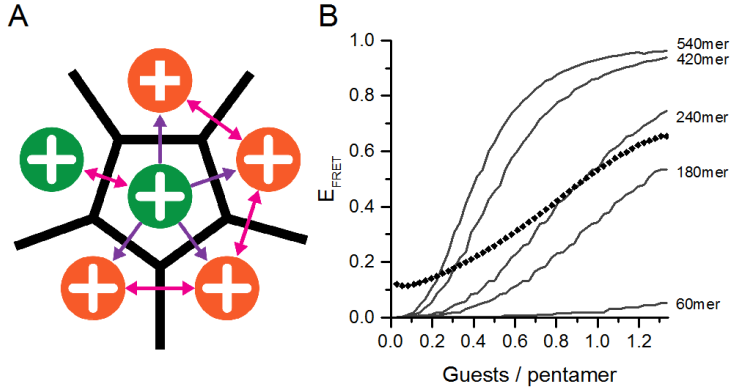


Figure 36. FRET within the AaLS-13 capsid. (A) Schematic representation of one GFP36+ donor (green) surrounded by 4 TOP36+ acceptors (orange) as FRET acceptors (purple arrow) and one GFP36+ homo-FRET partner (magenta arrow). (B) FRET-efficiencies as a function of the guest:capsid protein ratio, measured (diamonds) and calculated for different oligomeric states given an inner diameter d = 26 nm.

3.2.2 Equilibrium properties of capsid-guest complexes

A concentrated sample of AaLS-13 capsids containing a mixture of GFP36+ and TOP36+ was diluted in buffers of different pH, and the emission spectra were measured (Figure 37A). Between pH 8.0 and 10.0, the ratio of emission at 525 and 508 nm remains constant, indicating a stable guest-host complex. The decrease in FRET observed between pH 10-11 most likely reflects reduced affinity of the guest proteins to the capsid due to deprotonation of lysines in the positively supercharged guest proteins. Above pH 11.0 the fluorophores are slowly degraded. The decrease in FRET below pH 8.0 is more likely attributable to reduced fluorescence of TOP36+ relative to GFP36+ than guest release from the capsid, since the increased chromophore pK_a of 6.6 for YFP versus 5.8 for GFP at 200 mM chloride is likely mirrored in the supercharged variants [188].

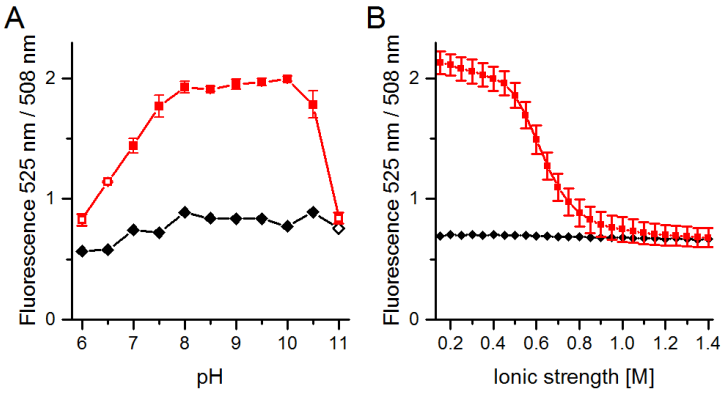


Figure 37. Relative fluorescence at 525 nm and 508 nm for a mixture of GFP36+/TOP36+ free in solution (black diamonds) or encapsulated by AaLS-13 (red squares) as a function of (A) pH and (B) ionic strength. Open symbols denote capsid precipitation (pH ≤ 6.5) or partial fluorophore degradation (pH ≥ 11.0). Error bars indicate the standard deviation from three independently prepared capsid-cargo complexes with 35 ± 3 guests per 180-mer capsid and a TOP36+:GFP36+ ratio of 4.0 ± 0.7.

Titrating concentrated aqueous NaCl into a solution of AaLS-13 capsids filled with a 1:4 mixture of GFP36+ and TOP36+ results in a decrease in TOP36+ emission and an increase in GFP36+ emission, reflecting decreasing FRET efficiency (Figure 37B). In order to exclude halide-dependent fluorescence quenching as the source of the decrease, the titration experiment was repeated with Na_2SO_4, which gave analogous results (Figure 38). Ionic strength-dependent release of guest proteins from capsids was confirmed by size-exclusion chromatography between I = 150 mM and 1,150 mM (Figure 39). TEM of the capsid-containing fractions shows exclusively intact capsids (Figure 40), suggesting that AaLS-13 releases guest proteins without permanently compromising capsid integrity.

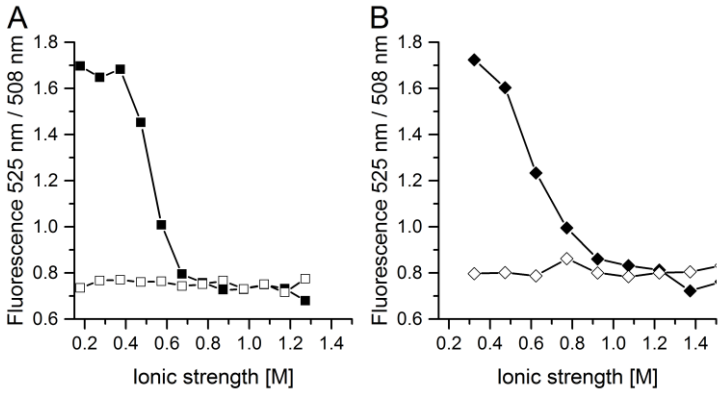


Figure 38. FRET between GFP36+ and TOP36+ in the absence (empty symbols) or presence of AaLS-13 capsid (full symbols) as a function of ionic strength. (A) Ionic strength modulated by addition of NaCl. (B) Ionic strength modulated by addition of Na_2SO_4. Note: Ionic strength is calculated by $I = 0.5 \sum_{i=1}^{n} c_i z_i^2$. As is apparent from the different inflection points for NaCl and Na_2SO_4, the Coulombic energy does not scale with the ionic strength, but rather the Debye length κ^{-1} which is given by $\kappa^{-1} = \frac{0.304\ nm}{\sqrt{[NaCl]/M}}$ or $\kappa^{-1} = \frac{0.176\ nm}{\sqrt{[Na_2SO_4]/M}}$ [196].

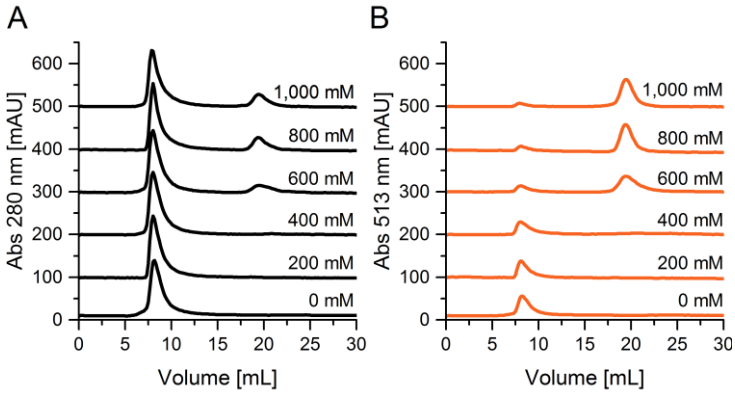


Figure 39. Size-exclusion chromatograms of AaLS-13 capsid containing a 1:4 mixture of GFP36+/TOP36+ at different ionic strengths (50 mM NaPi, 5 mM EDTA, pH 8.0 and varying concentrations of NaCl). (A) Absorption at 280 nm for AaLS-13 and guests. (B) Absorption at 513 nm for GFP36+ and TOP36+. Note: below 400 mM NaCl, free GFP36+ and TOP36+ adhere to the stationary phase (cross-linked agarose).

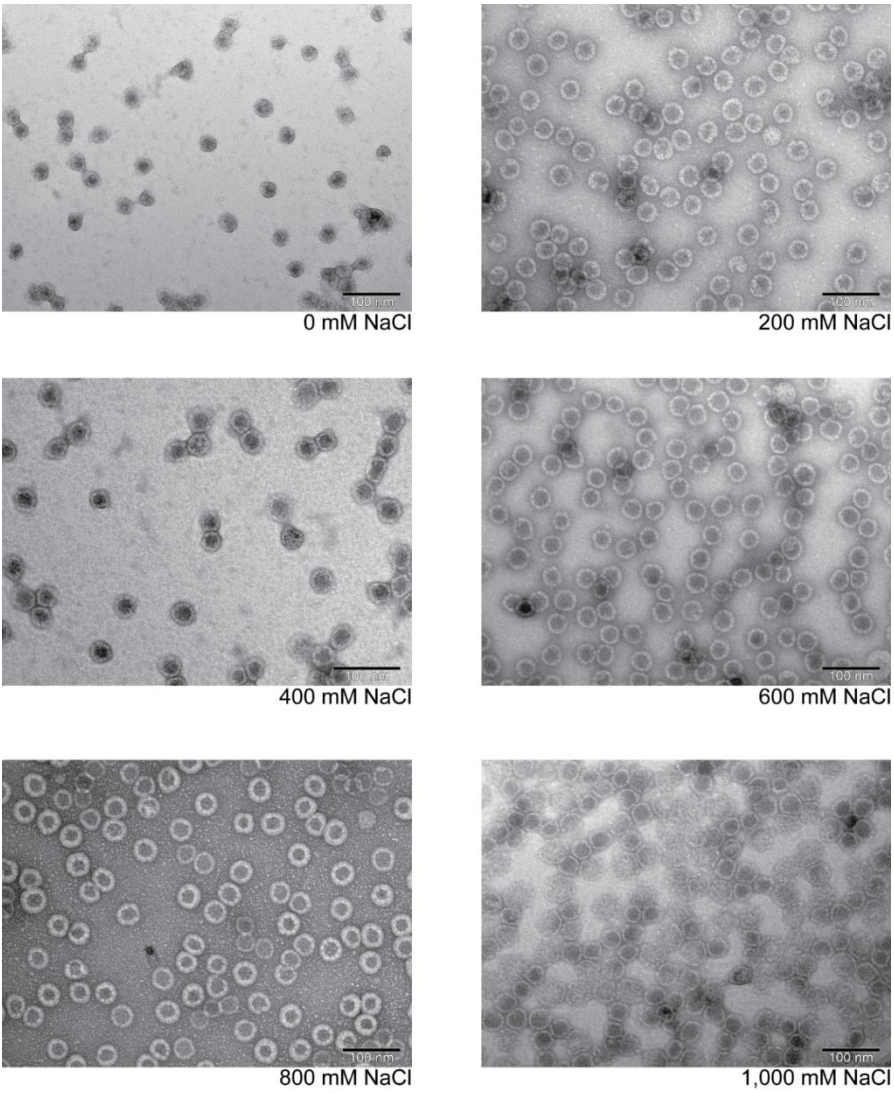


Figure 40. Negative-staining transmission electron microscopy pictures of AaLS-13 capsids at different ionic strengths (50 mM NaP_i, 5 mM EDTA, pH 8.0 and varying concentrations of NaCl).

In order to determine whether cargo molecules are released from AaLS-13 capsids even at low ionic strength, capsids filled with GFP36+ and capsids filled with

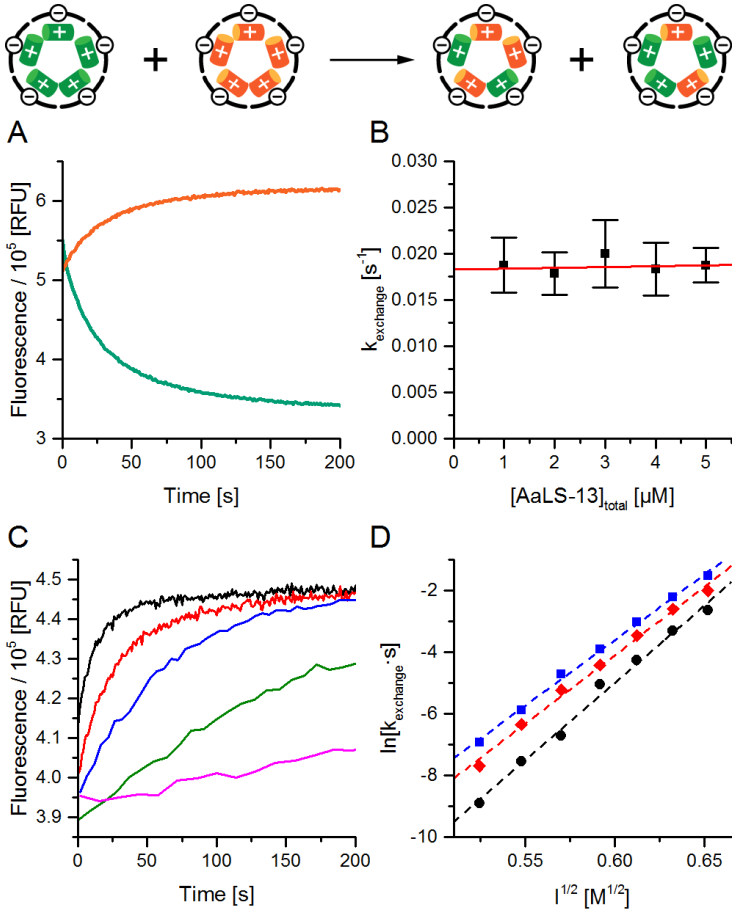


Figure 41. Guest exchange between filled AaLS-13 capsids. (A) Time dependent emission at 508 nm (green) and 525 nm (orange) upon mixing capsids containing GFP36+ with capsids containing TOP36+ (32 guests/180-mer capsid, I = 350 mM). (B) Rate of guest exchange as a function of capsid concentration (32 guests/capsid, I = 350 mM). To a significance level <0.05, the exchange is not dependent on AaLS-13 concentration. (C) Traces of 525 nm emission for guest exchange between AaLS-13 capsids containing 22 guests/capsid at ionic strengths 425 mM (black), 400 mM (red), 375 mM (blue), 350 mM (green) and 325 mM (magenta). (D) Rate of guest exchange as a function of the square root of ionic strength for AaLS-13 capsids containing 22 guests/capsid (black circles), 28 guests/capsid (red diamonds) and 32 guests/capsid (blue squares).

TOP36+ were mixed and the change in fluorescence emission recorded (Figure 41A). Both the increase in emission at 525 nm and decrease at 508 nm follow an

exponential time-dependence, indicating rapid scrambling of cargo proteins between capsids. The half-life of this process is 37 ± 4 s for 32 guests/180-mer capsid at an ionic strength of 350 mM. Exchange of guests could occur by three distinct mechanisms: i) guests are transferred upon collision of at least two capsids; ii) guest molecules dissociate from one capsid and enter another; or iii) capsid fragments detach with bound cargo and invade another capsid shell.

Figure 41B demonstrates that the rate of guest exchange is independent of capsid concentration, which excludes mechanisms in which the rate-determining step involves more than one capsid. Dissociation of a capsid subunit and subsequent incorporation into another capsid appears unlikely based on an experiment in which empty capsids were covalently labeled with a fluorescent dye (Figure 42 and Figure 43). No FRET was observed within 24 hours of mixing AaLS-13 capsids labeled with ATTO-495 and capsids labeled with ATTO-565 (a FRET pair). Capsids assembled from fragments that were respectively labeled with the two fluorophores served as a positive control.

Having excluded both a bimolecular mechanism and the exchange of capsid fragments, transient dissociation of guest molecules from the capsid is the most likely explanation for guest exchange. The logarithm of the rate of exchange is linearly dependent on the square root of the ionic strength (Figure 41C,D), and the corresponding free energy of activation follows the same dependency. Since the free energy for Coulombic interactions in aqueous salt solutions scales with the square root of ionic strength[196,197], overcoming the electrostatic attraction between positively charged guest proteins and the negatively charged capsid shell is likely the rate-determining step. With increasing occupancy, the electrostatic binding energy per guest decreases as the attraction between the guest and the lumenal capsid surface is somewhat offset by electrostatic repulsion between guests, resulting in a faster guest exchange rate for more packed capsids (Figure 41D). This effect becomes less pronounced at higher ionic strength, which shields the charges on the protein surfaces more effectively.

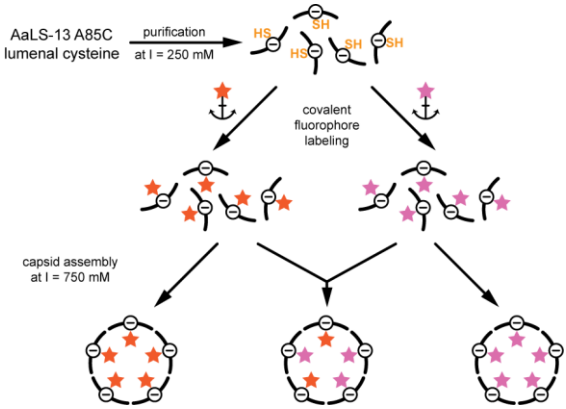


Figure 42. Strategy for obtaining homo- and heterogeneously covalently labeled AaLS-13 A85C capsids. Purification by metal affinity chromatography from cell lysate at an ionic strength of 250 mM afforded exclusively pentameric capsid fragments. These fragments were covalently labeled with either ATTO495-maleimide or ATTO565-maleimide. Increasing the ionic strength to 750 mM gave homogeneously labeled assembled capsids after one week at room temperature. Incubating a 1:1 mixture of AaLS-13 A85C(ATTO495) and AaLS-13 A85C(ATTO565) yielded capsids containing both fluorescent labels, which served as a positive control for subsequent experiments.

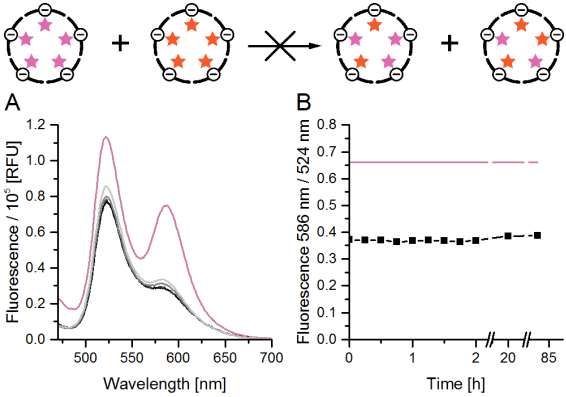


Figure 43. Exchange of AaLS-13 subunits between assembled capsids. (A) Emission spectra of AaLS-13 capsid labeled with both ATTO495 and ATTO565 (pink) or a mixture of AaLS-13 capsids labeled with either ATTO495 or ATTO565 after 1 min (black), 100 min (dark grey), 1 day (grey) and 4 days (light grey) upon excitation at 450 nm. (B) Time dependence of the ratio of ATTO565 (586 nm) to ATTO495 (524 nm) emission upon excitation at 450 nm after mixing AaLS-13 capsids labeled with either ATTO495 or ATTO565 (black squares). The ratio for AaLS-13 capsids containing both fluorophores is shown as a pink line.

3.2.3 Kinetics of guest encapsulation

Since the release of guest molecules from AaLS-13 capsids is clearly dominated by the strength of Coulombic interactions, we became interested in the rate of encapsulation. Previously, mixtures of GFP36+ and AaLS-13 capsids were incubated overnight before analysis by size-exclusion chromatography [181]. However, preliminary kinetic experiments showed that the fluorescence changes attributed to co-encapsulation of GFP36+ and TOP36+ were already complete within a few seconds. We therefore monitored cargo loading by stopped-flow spectroscopy.

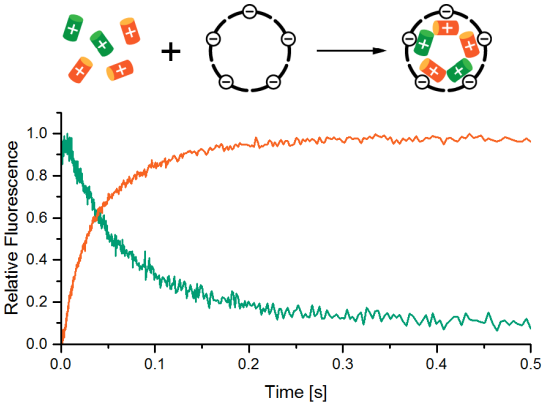


Figure 44. Fluorescence emission at 500 ± 5 nm (green) and >515 nm (orange) after mixing 20 µM empty AaLS-13 capsids with 4 µM 1:4 GFP36+/TOP36+ at I = 550 mM in a stopped flow apparatus.

Guest molecules (GFP36+ and TOP36+ in a 1:4 ratio) and empty AaLS-13 capsids were mixed in a stopped-flow device at an ionic strength of 550 mM and fluorescence upon selective excitation of GFP36+ at 400 nm was followed. The time course of GFP36+ fluorescence was recorded using a band pass filter with transmission between 495-505 nm; TOP36+ fluorescence was recorded using a long pass filter with transmission above 515 nm. Consistent with results presented in Figure 35, encapsulation increases FRET efficiency as apparent from a decrease in GFP36+ donor emission and an increase in TOP36+ acceptor emission (Figure 44). Surprisingly, however, encapsulation as indicated by intermolecular FRET between guest molecules was complete within 400 ms. Given the rigidity generally attributed to capsid shells [167,169] and failure to observe exchange of subunits between AaLS-13 capsids (Figure 43), rapid uptake of proteins as large as 30 kDa is startling. Consequently, several control experiments were performed to exclude spectroscopic artefacts.

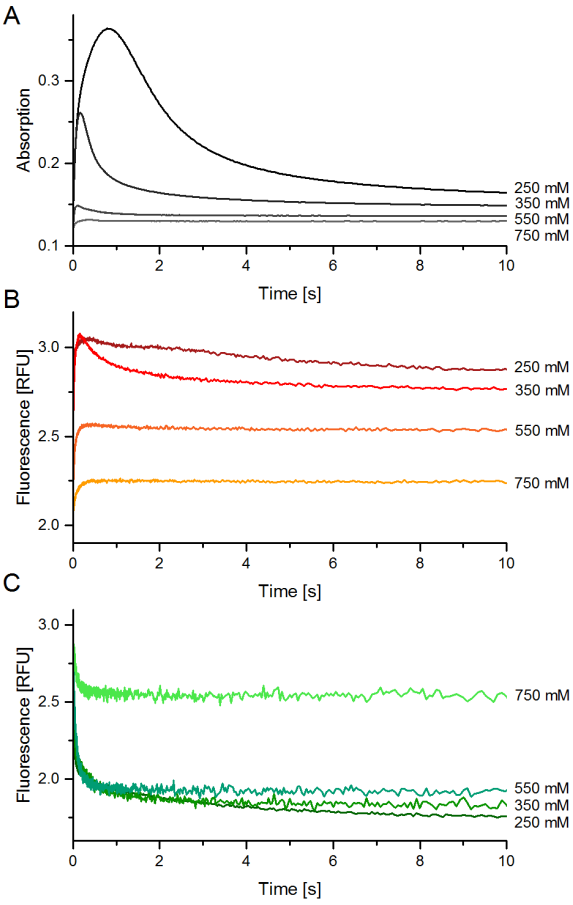


Figure 45. Progress curves after mixing 20 µM AaLS-13 capsids with 4 µM guest at different ionic strengths. Protein concentrations and instrument parameters were kept constant, allowing quantitative comparison. (A) Light scattering measured by absorption at 400 nm. (B) Emission > 515 nm. (C) Emission 500 ± 5 nm.

When supercharged guest molecules were mixed with AaLS-13 capsids on a preparative scale, the solutions became instantly turbid but clarified within seconds. Turbidity is a result of an interaction between light and particles with diameters similar to the wavelength of the observed light. It can be detected spectroscopically by an increase in absorption proportional to λ^{-4} or by light scattering quantified by a photomultiplier positioned at a 90° angle relative to the incident beam. Both tech-

niques were employed in the stopped flow setup to monitor the transient formation of the strongly scattering species generated upon mixing guests with capsids (Figure 45A).

The ionic strength at which guests and capsids are mixed has a profound impact on the progress curves observed by stopped flow spectroscopy (Figure 45). The decreasing amplitudes for changes in TOP36+ and GFP36+ emission upon increasing ionic strength reflect the reduced degree of encapsulation at higher salt concentrations. At low ionic strength (I = 250 mM), an immediate and pronounced increase in light scattering is observed, which relaxes within a few seconds. Simultaneously, the TOP36+ emission reaches a local maximum within a few hundred milliseconds and subsequently relaxes (Figure 45B). In contrast, GFP36+ emission decreases continuously in biphasic fashion (Figure 45C). The multi-phasic fluorescence changes and the transient increase in light scattering indicate the existence of at least one intermediate during encapsulation.

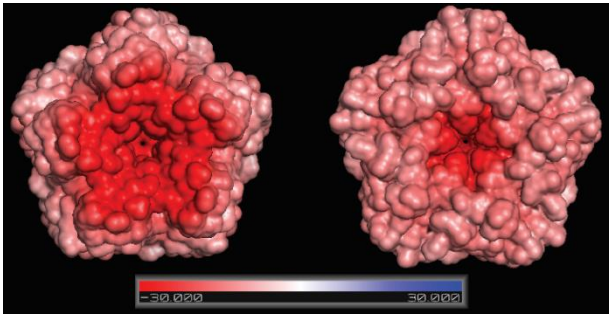


Figure 46. Electrostatic surface potential of the interior (left) and exterior surface (right) of a AaLS-13 pentamer at I = 150 mM. The color gradient depicts the potential within ± 30 k_BT/e.

This intermediate likely involves supercharged guests binding to the capsid exterior. Calculations of the surface electric potential of an AaLS-13 pentamer revealed substantial negative charge on the outer surface (Figure 46). Using the APBS software [53], the potential on the exterior face of the pentamer was calculated to be about two thirds that of the interior face (-14 k_BT/e and -24 k_BT/e, respectively, at I = 150 mM). Upon first encountering empty capsids, the positively charged guest molecules would initially bind to the negatively charged capsid exterior. Crossing the capsid shell and entering the lumen should then be limited, at least partially, by dissociation from the surface. In order to detach, the guests need to overcome the electrostatic attraction to the outer capsid wall, the strength of which is strongly

dependent on the concentration of screening ions. In accordance with this hypothesis, the signals attributed to the aggregate—TOP36+ fluorescence and light scattering—are much less pronounced at higher ionic strength (Figure 45), indicating faster decay of the intermediate.

Although decay of the intermediate is likely a unimolecular process, its formation must be at least bimolecular and should therefore be concentration dependent. Indeed, when AaLS-13 capsids of different concentration are mixed with guests at a constant stoichiometric ratio, the signals associated with the intermediate become less intense with decreasing concentration, until below 7 µM AaLS-13 both signals show a continuous approach to equilibrium rather than a transient maximum. (Figure 47A).

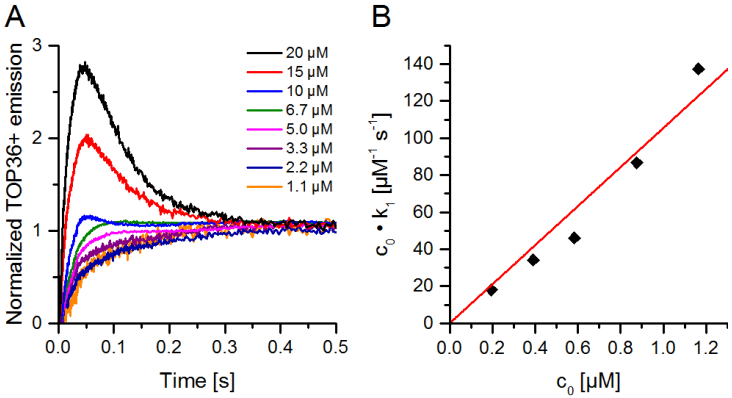


Figure 47. Concentration dependent guest encapsulation by AaLS-13 capsids. (A) Progress curves for TOP36+ emission upon mixing 1:4 mixtures of GFP36+/TOP36+ with AaLS-13 at different protein concentrations (concentrations in legend with respect to AaLS-13 monomer). (B) Apparent rate constant for guest encapsulation as a function of initial protein concentration at an ionic strength of 350 mM.

Quantitative analysis of these progress curves is tempting but has to be undertaken with caution. Due to low FRET efficiency at low loading densities and precipitation induced by of excess guest, only a narrow range of capsid:guest ratios can be probed, precluding experiments under pseudo-first-order conditions. Since pseudo-first-order experiments are impractical, association rates were estimated assuming second-order reactions between guests and capsids and fitting the progress curve to the corresponding rate law. Regression of the apparent rate constant against the initial concentrations of guests and capsids gave a bimolecular rate constant for the encapsulation of GFP36+/TOP36+ mixtures of 10^8 M^{-1} s^{-1} at 350 mM

ionic strength, pH 8.0, 25 °C (Figure 47B). From the values of k_{on} and k_{off}, the dissociation constant for the GFP36+/capsid complex is estimated to be about 0.2 nM. The effect of salt concentration on binding affinity can be approximated by the dependence of the rate constant for guest exchange on ionic strength (Figure 41D), assuming that k_{on} is not significantly affected. The predicted stabilities agree well with the values observed by FRET (Figure 48).

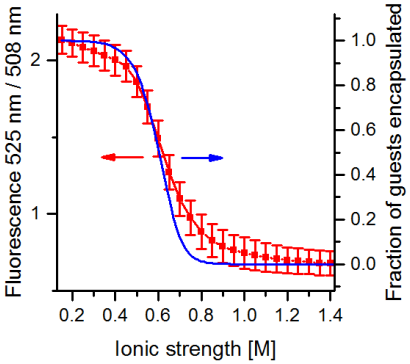


Figure 48. Comparison of the ionic strength dependence of guest encapsulation by AaLS-13 measured by FRET and calculated from k_{on} and k_{off}. The experimentally determined relative fluorescence at 525 nm and 508 nm for a mixture of GFP36+/TOP36+ indicates the extent of encapsulation as a function of ionic strength (red squares; data are identical to Figure 37B). The fraction of guests encapsulated at 1 µM guest and 5 µM AaLS-13 (concentration with respect to monomer) was calculated from the K_D value and is shown as a blue line. The dissociation constant for binding of GFP36+ or TOP36+ to AaLS-13 was estimated using the equation $K_D = k_{off}/k_{on}$ and the rate constants $k_{on} \approx 10^8 \mathrm{M}^{-1}\mathrm{s}^{-1}$ and $k_{off}(I) \approx e^{a+b\cdot\sqrt{I}}$. The parameters a = -29.1 ± 0.6 and b = 42.5 ± 1.0 $\mathrm{M}^{-1/2}$ were determined by a linear fit of the logarithmized rate of guest exchange versus the square root of ionic strength for an AaLS-13 capsid containing 32 guests per 180-mer capsid (Figure 41D).

3.2.4 Probing the encapsulation mechanism with fluorescently labeled AaLS-13 variants

Existence of an intermediary complex with surface-bound guests rationalizes the transient increase in TOP36+ fluorescence, which is likely more environmentally sensitive than GFP36+, assuming that the trend observed for non-supercharged GFP variants holds [188,198]. In order to obtain further evidence regarding the nature of this intermediate, we designed two capsid variants that were labeled with the FRET acceptor ATTO-565 on either the interior or exterior face of the capsid shell. If the intermediate has supercharged guests bound to the capsid exterior,

FRET between TOP36+ and the fluorophore on the outer capsid surface should decrease simultaneously with the decay of this intermediate. The opposite trend should apply for capsids labeled on the lumenal face.

Since AaLS-13 contains three nucleophilic thiols per monomer, we created a variant named AaLS-RR (AaLS-13 C52R/C127R) by reverting two surface-exposed cysteine residues, which had appeared during directed evolution, to arginine as found in wild-type AaLS. The third thiol, Cys37, is buried and its reactivity towards electrophilic labeling reagents is therefore expected to be negligible, particularly when substoichiometric amounts of fluorophore reagent are employed. Introduction of exposed cysteines into this variant afforded capsids that can be selectively labeled with ATTO-565-maleimide on either the interior (A85C) or exterior (R108C) surface [199].

AaLS-RR is less negatively charged than AaLS-13 (because of reintroduction of two cationic arginines) and forms slightly smaller capsids with an average outer diameter of 32 ± 2 nm. In contrast to AaLS-13, capsid formation is reversible at an ionic strength of 350 mM. When purified AaLS-RR capsids are analyzed by size-exclusion chromatography, an approximately 1:1 mass ratio of assembled capsid to capsid fragments is obtained. Nevertheless, the changes in fluorescence upon encapsulation are qualitatively similar to those observed with AaLS-13 (Figure 49A). The life-time of the transient intermediate is approximately threefold reduced compared to AaLS-13, presumably due to weaker Coulombic interactions between TOP36+ and the less charged AaLS-RR capsid. The most notable difference is the increase in both light scattering and TOP36+ fluorescence after 0.2 seconds, which was not observed for AaLS-13 (Figure 45), and probably reflects capsid formation from free capsid fragments triggered by binding of supercharged guest proteins. Although the AaLS-13 and AaLS-RR capsids are less alike than expected, the similar spectroscopic trends observed (Figure 45 and Figure 49A) suggest that both capsid variants have analogous loading mechanisms.

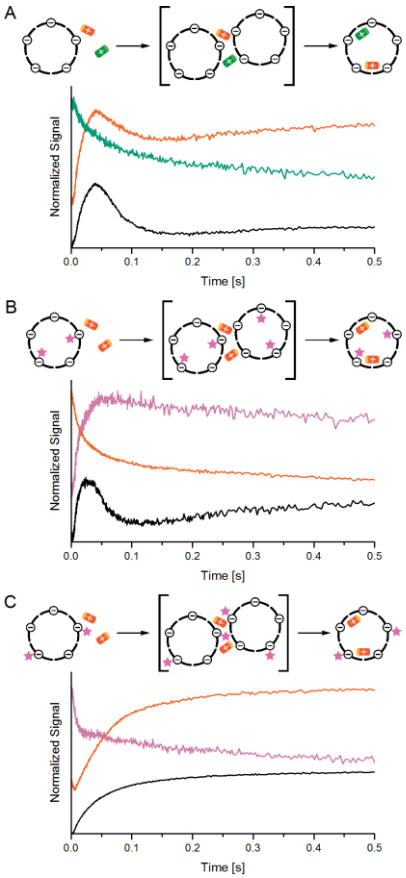


Figure 49. Change in GFP36+ emission (green), TOP36+ emission (orange), ATTO-565 emission (pink) and light scattering (black) after mixing 20 µM AaLS-RR variants with 4 µM guest proteins at I = 350 mM. (A) Encapsulation of a GFP36+/TOP36+ 1:4 mixture into AaLS-RR. (B) Encapsulation of TOP36+ into AaLS-RR*^{IN} labeled on the lumenal face. (C**)** Encapsulation of TOP36+ into AaLS-RR*^{OUT} labeled on the exterior face.

The time course of the fluorescent signals observed upon mixing covalently labeled AaLS-RR capsids with TOP36+ indicate an initial interaction between TOP36+ and the outside of the shell. Towards completion of the encapsulation process by capsids with an interior label (AaLS-RR*^{IN}), FRET increases as the distance between Top36+ and ATTO-565 decreases, observable by a decrease in donor (TOP36+) fluorescence and an increase in acceptor (ATTO-565) emission (Figure 49B). Complementarily, when TOP36+ migrates from the external capsid surface to the interior,

the donor-acceptor distance increases for capsids labeled on the exterior (AaLS-RR*^{OUT}), resulting in rising donor (TOP36+) and decreasing acceptor (ATTO-565) emission (Figure 49C). Although mixing AaLS-RR*^{OUT} and TOP36+ ultimately leads to precipitation as apparent from the persistently strong light scattering, the initial fluorescence changes are consistent with migration of TOP36+ from the capsid exterior to the lumen.

3.2.5 Encapsulation by the AaLS-neg capsid

The kinetic experiments with AaLS-13 raise questions about the molecular basis for rapid cargo loading. In order to identify which mutations might confer these unusual properties, we investigated encapsulation by AaLS-neg, the engineered capsid that served as the starting point for directed evolution. Stopped-flow experiments showed that the time-scales for encapsulation of the GFP36+/Top36+ FRET pair are comparable to those for AaLS-13 and AaLS-RR, albeit with a diminished propensity to form intermediate aggregates (Figure 50A). The half-life for cargo exchange between filled AaLS-neg capsids was determined to be 13 ± 7 s—about threefold shorter than for AaLS-13 (Figure 50B). These results suggest that the key features enabling capsids to take up globular proteins are already present in AaLS-neg, and must therefore be a direct consequence of introducing additional negative charge on the lumenal surface of the capsid.

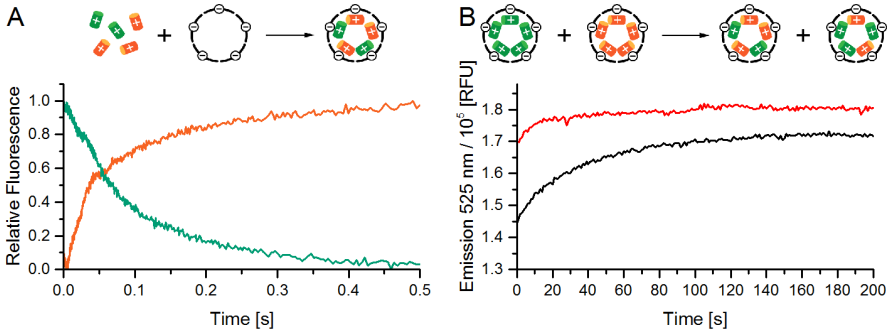


Figure 50. Kinetics of encapsulation and guest exchange for AaLS-neg. A) Fluorescence emission at 500 ± 5 nm (green) and >515 nm (orange) after mixing 20 µM empty AaLS-neg capsids with 4 µM 1:4 GFP36+/TOP36+ at I = 350 mM in a stopped flow apparatus. B) Time dependent emission at 525 nm upon mixing AaLS-neg (red) or AaLS-13 (black) capsids containing GFP36+ with capsids containing TOP36+ (30 guests/180-mer capsid, I = 350 mM). The different endpoints reflect higher FRET efficiencies likely caused by denser loading of the smaller AaLS-neg capsids (diameter = 28.6 ± 2.6 nm versus 35.4 ± 3.2 nm for AaLS-13).

3.2.6 A kinetic model for guest encapsulation

Encapsulation of supercharged guests into assembled AaLS-13 capsid is a robust, reversible process and is unexpectedly fast. Mixing guest and capsid does not immediately afford loaded capsids, but most likely proceeds through an intermediate that exhibits pronounced light scattering. Stopped-flow experiments with TOP36+ and inside or outside labeled AaLS-13 capsids suggest that guest molecules are transiently associated with the exterior of the capsids in the intermediate. Based on these observations we propose a 2-step kinetic model for the encapsulation of supercharged guests by AaLS-13 capsids (Figure 51).

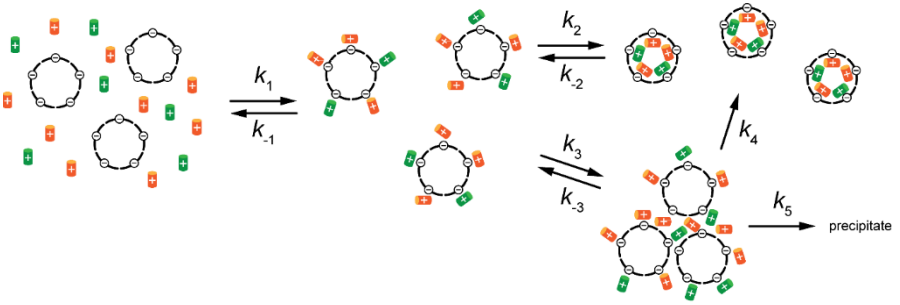


Figure 51. Kinetic model for guest encapsulation in AaLS-13 capsids.

Upon encountering the negatively charged shells, the positively charged guests first bind weakly to the exterior capsid surface (k_1). The strength of this interaction, modulated by the ionic strength, determines the rate at which the surface bound guest can dissociate and enter the capsid (k_2). If this transition is slow, the surface bound guest molecules can act as bridges between capsids and promote aggregation (k_3). The resulting colloids cause a transient increase in light scattering. If the interactions are sufficiently weak, the supercharged guest proteins internalize, gradually increasing the negative charge on the capsid's external surface until the individual particles repel each other and the colloids disintegrate (k_4). If, however, the Coulombic interactions between guest and capsid surface are too strong, the colloids will continue to grow and eventually precipitate (k_5). For example, mixing 20 µM empty AaLS-13 capsids with 8 µM GFP36+/TOP36+ leads to precipitation at an ionic strength of 150 mM. A similar mechanism has been proposed by Malyutin and Dragnea to explain intermediate aggregation during nanoparticle-templated assembly of Brome mosaic virus, albeit with opposite roles for capsid protein and

guest [200]. In their model, individual capsid proteins act as bridges between nanoparticles and the dissolution of the colloid is driven by the budding of completed capsids.

Whether mixing empty capsids with cargo results in encapsulation or precipitation depends on the relative rates of the two processes, which can be tuned by varying ionic strength, protein concentration or charge on the guest molecules. For GFP36+, the previously established standard conditions (50 mM NaP_i, 200 mM NaCl, 5 mM EDTA, I = 350 mM)[181] are a good compromise between low salt concentration, which favors capsid-cargo complex formation thermodynamically but leads to unwanted precipitation, and high salt concentration, which diminishes precipitation but also reduces encapsulation yields. For more positively charged oligomeric guests, precipitation can be avoided by increasing ionic strength during encapsulation [184]. The resulting complexes can then be stabilized by subsequently reducing the salt concentration.

3.2.7 Conclusions from kinetic experiments

By virtue of multivalent interactions between constituent monomers, viral capsids are extremely stable structures [167]. In the absence of an appropriate trigger, dissociation of subunits occurs very slowly, if at all [201]. Similarly, the subunit exchange experiments indicate high kinetic stability of assembled AaLS-13. In this light, rapid internalization of 30 kDa proteins into assembled AaLS-13 capsids is rather surprising and highlights a fundamental difference between this engineered system and typical viral capsids. Applying established principles used to study protein-protein interactions, a bimolecular rate constant of 10^8 M^{-1} s^{-1} could be determined for guest encapsulation, which approaches the diffusion limit. Since the internalization rate depends on protein concentration, the rate-limiting step must involve the encounter of guests and capsids rather than intrinsic motions of the capsid. Consequently, the protein motions that allow passage of a 30 kDa protein are either surprisingly fast ($\tau < 5$ ms), or the responsible gateways are already in place.

Rapid encapsulation goes hand in hand with facile release of guest proteins from AaLS-13 capsids. Under standard conditions (I = 350 mM), GFP36+ molecules exchange between capsids with a half-life of less than one minute. In contrast, under physiological conditions (I = 160 mM) the half-life increases to 36 hours. The latter result implies that on timescales relevant for in vivo applications, the capsid-cargo complex can be regarded as kinetically stable. This property is desirable for drug delivery applications in which the guests have to remain shielded before reaching their designated destination.

The experiments demonstrate that the AaLS-13 capsid shell does not pose a physical barrier to molecules as large as GFP, implying that engineering and evolution of AaLS-13 to accommodate positively charged cargo must have had a dramatic effect on its structure, dynamics, or both. Since wild-type AaLS contains pores with a diameter of only 9 Å, the mutations incorporated during the design and subsequent directed evolution of AaLS might have either destabilized the interactions between subunits or endowed the shell with voids large enough to allow transport of sizeable proteins across the capsid wall.

Alternative explanations, for example that the endpoint of the time course for FRET acceptor emission corresponds to guest molecules bound to the exterior surface of the capsid, can be excluded based on internalization experiments with capsids that were labeled with fluorophores on either the interior or exterior surface. Pre-organized voids in the capsid shell could hint at incomplete protein shells or might be an inherent feature of the evolved AaLS-13 capsid. Complete capsids bearing sizeable holes are not unprecedented. Bacteriophage P22, for instance, releases a pentamer upon incubation at 70 °C in vitro, yielding a stable capsid with a pore spanning 10 nm [202].

The notion that the AaLS-13 capsid is highly flexible is supported by the observation that reversion of two mutations, C52R and C127R, afforded capsids that are kinetically labile at low ionic strength (≤ 350 mM). Hence, subunit interactions in AaLS-13 might be just strong enough to prevent detectable dissociation within a few days, but sufficiently weak to generate transient clefts that allow passage of globular proteins.

Judging from the similar rates of guest encapsulation and cargo exchange for the engineered AaLS-neg and subsequently evolved AaLS-13 variants, the structural features that allow facile entry of protein cargo into assembled capsids were already present in the original design. Electrostatic repulsion between the additional negative charges introduced on the lumenal face of the shell protein may have rendered the capsid subunits less compact and reduced their Gaussian curvature. Adaptation of the subunit structure to the increased local charge density in this way might have enabled formation of higher-order polyhedral capsids (T=3, 4 or 7) and at the same time introduced the flexibility needed for globular proteins to cross the shell.

Since the ability to encapsulate cargo by assembled capsids is already inherent to AaLS-neg, the improved sequestration achieved by directed evolution presumably reflects the increased charge density, resulting in tighter binding of positively charged guests [179]. Given the unexpected destabilization of the AaLS-13 capsid upon simple reversion of two residues, it is unlikely that rational design would have

identified the optimal compromise required to maximize negative surface charge and maintain capsid stability.

3.3 Probing capsid formation in vivo

When AaLS-13 is purified by metal affinity chromatography in 50 mM NaP_i, 300 mM NaCl, pH 8.0, a mixture of assembled capsid and pentameric subunits is obtained. In contrast, purification at lower ionic strengths (e.g., 50 mM NaP_i, 100 mM NaCl, pH 8.0; I = 250 mM) affords exclusively fragments and no capsids—a phenomenon that is exploited for assembling capsids consisting of subunits with different fluorescent labels (Figure 42). Failure to obtain capsids at an ionic strength of 250 mM raises the question whether AaLS-13 can assemble into capsids at lower ionic strength, particularly in the cytosol of *E. coli* (I ≈ 150 mM). Conceivably, AaLS-13 capsids might not form in the cytosol and, instead, only assemble post-lysis by virtue of the increased ionic strength of the lysis buffer.

In order to test whether AaLS-13 forms capsids in vivo, the GFP36+/TOP36+ FRET pair was employed as a probe. The two fluorescent proteins were co-expressed in *E. coli* KA13 [203] under control of the tetracycline repressor and different AaLS variants were concomitantly expressed using the T7 promoter under control of the *lac* operon (Figure 52). Only if AaLS-13 forms capsids in vivo, would one observe FRET between the two fluorescent proteins. AaLS-neg, which is always and exclusively obtained as assembled capsids are after purification, served as a positive control. AaLS-wt, which is not known to encapsulate positively supercharged cargo molecules, and AaLS-QQ, a variant with the same overall charge as AaLS-13 but unable to form capsids, were employed as negative controls.

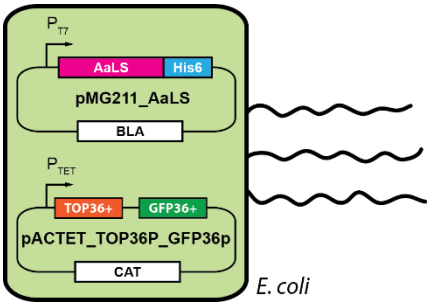


Figure 52. Two-plasmid system to co-express TOP36+, GFP36+ and an AaLS variant.

Cultures were grown in the presence of different concentrations of tetracycline and, after 5 hours, induced with IPTG. The ratio of GFP36+/TOP36p to AaLS is dependent on the tetracycline concentration (Figure 53A). At high tetracycline concentrations, growth of the cells is severely inhibited, as apparent from growth curves (Figure 53B), resulting in reduced amounts of protein.

The different normalized emission spectra recorded upon irradiation at 420 nm, which selectively excites GFP36+ over TOP36+, for AaLS-wt and AaLS-neg, and AaLS-13 and AaLS-QQ respectively (Figure 53C), are probably not attributable to different extents of GFP36+/TOP36+ co-encapsulation. As apparent from Figure 53D, which shows absolute fluorescence at 508 nm (filled symbols) and 525 nm (open symbols), the fluorescence attributed to TOP36+ significantly increased after induction of AaLS-wt and AaLS-neg, but not AaLS-13 and AaLS-QQ expression. Both AaLS-wt and AaLS-neg bind flavins and related molecules that exhibit similar fluorescence as TOP36+ [204], whereas this ability has been lost in AaLS-13 and AaLS-QQ. It is conceivable that binding at the hydrophobic pockets of AaLS-wt and AaLS-neg enhances flavin fluorescence. Alternatively, over-expression of lumazine synthase could increase the overall flavin concentration in the culture. It is not known whether AaLS-neg is still capable of catalyzing lumazine synthesis. Due to this interference, the GFP36+/TOP36+ FRET pair cannot be employed as a probe to determine the assembly state of AaLS-13 in vivo.

Protein complexes obtained after co-expressing GFP36+, TOP36+ and AaLS variants in *E. coli* KA13 were purified at low ionic strength (50 mM TRIS, 100 mM NaCl, pH 8.0) by metal affinity chromatography. GFP36+, TOP36+, both lacking the His6-tag, could only be co-purified by binding to AaLS-variants bearing a C-terminal His6-tag. As evident from the SDS-PAGE analysis of purified complexes, GFP36+/TOP36+ were only co-purified with AaLS-neg, AaLS-13 and AaLS-QQ, but not AaLS-wt (Figure 54). The size-exclusion chromatograms indicate that the guests are encapsulated in intact capsids in the case of AaLS-neg. For AaLS-QQ on the other hand, no capsid formation was observed as expected. Nonetheless, the fluorescent proteins still remained associated with the shell protein fragments throughout the entire purification, which included multiple washing steps. Lastly, AaLS-13 elutes as a mixture of quaternary states (Figure 55), namely capsid together with some fluorescent protein, a species of intermediate size with which most of the fluorescent protein is associated, and fragments. Lack of fluorescence in the fractions eluting at the volume expected for pentamers (ca. 17 mL) suggests that binding of GFP36+ to AaLS-13 fragments induces formation of a higher-order structure, potentially an intermediate in capsid assembly. It elutes at approximately the same volume as AaLS-wt, and could therefore correspond to a hexamer of pentamers, which equates to an AaLS-wt half-sphere.

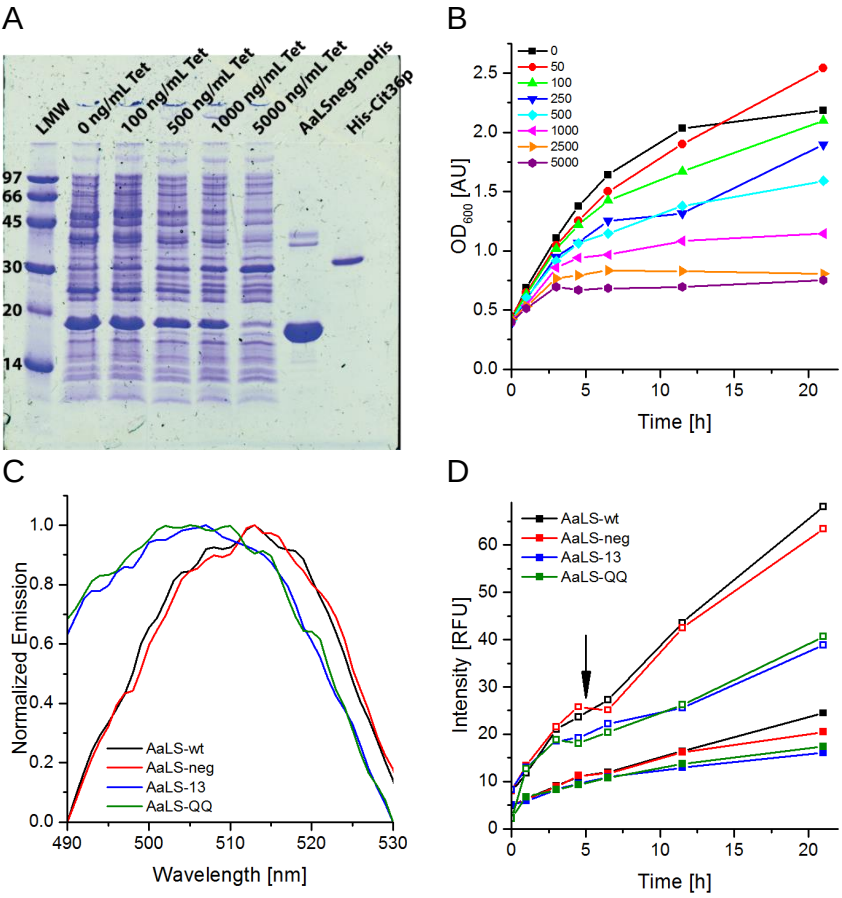


Figure 53. Co-expression of GFP36+/TOP36+ with different AaLS variants. A) SDS-PAGE analysis of protein production for the combination GFP36+/TOP36+ and AaLS-wt as a function of tetracycline concentration. His6-Cit36+, a spectral variant of His6-GFP36+ with equivalent molecular weight, was used for size comparison. B) Optical density at 600 nm, which scales with cell density, for different tetracycline inducer concentrations (values in the inset are in ng/mL). C) Normalized emission spectra of whole cell cultures upon excitation at 420 nm after growing for 21 h in the presence of 500 ng/mL tetracycline. D) Fluorescence time course for cell cultures grown in the presence of 500 ng/mL tetracycline (filled symbols: excitation 470 nm, emission 508 nm, corresponding to GFP36+ fluorescence; open symbols: excitation 510 nm, emission 525 nm, corresponding to TOP36+ fluorescence). The arrow indicates the time at which AaLS expression was induced by addition of 100 µM IPTG.

The absence of such a structure in the chromatogram of AaLS-QQ suggests that this species is not a GFP36+ molecule sandwiched between two pentamers, as

there is no apparent reason why such a structure could not form with AaLS-QQ as well.

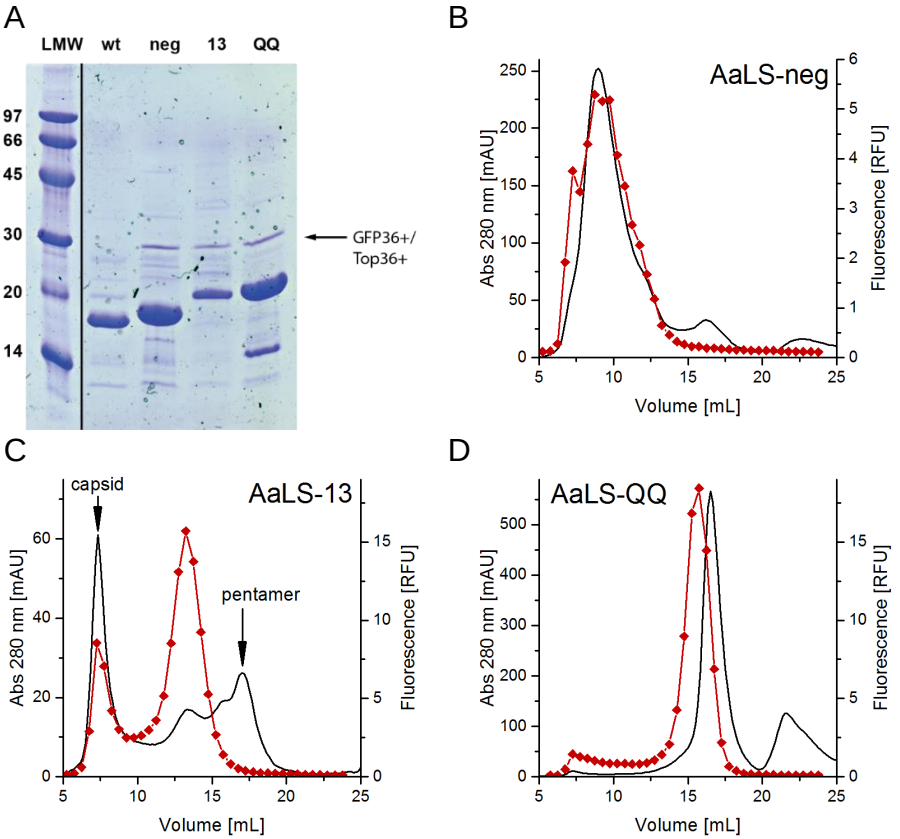


Figure 54. Co-purification of GFP36+/TOP36+ with AaLS variants. A) SDS-PAGE analysis of purified protein complexes for four different AaLS variants (the solid line indicates that the original gel contained additional lanes that were omitted here because they were not related to the experiment under discussion). B-D) Size-exclusion chromatograms of GFP36+/TOP36+ co-purified with the indicated AaLS variants. Black lines indicate on-line absorption at 280 nm, red lines indicate fluorescence of individual fractions (excitation 470 nm, emission 508 nm).

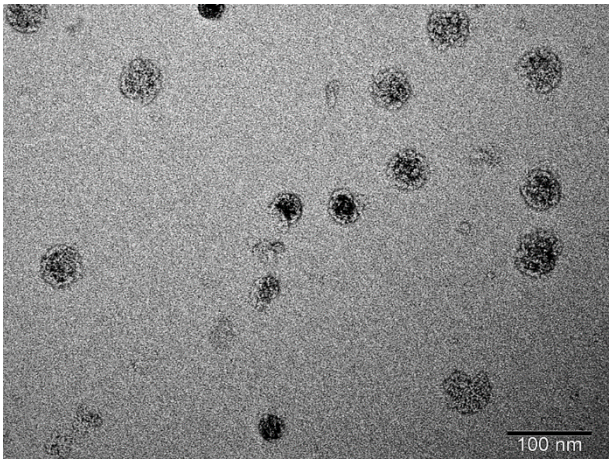


Figure 55. Electron microscopy picture of capsids purified at low ionic strength from bacterial cultures co-expressing GFP36+/TOP36+ with AaLS-13. The samples correspond to the peak marked "capsid" in Figure 54C. The contrast of this picture was enhanced.

A number of the AaLS variants that exhibited improved cell growth in the presence of R10-tagged HIV protease were found to contain frameshift mutations [179]. These constructs presumably reduced the toxicity of the enzyme by sequestering it in inclusion bodies. Along the same lines, one could speculate that the protective phenotype provided by co-expression of AaLS-13 and HIV protease may not depend on capsid formation. As an alternative explanation, the HIV protease dimer, equipped with two deca-arginine peptides (HIV-R10), might simply be deactivated by binding to a strongly negatively charged protein. To test this hypothesis, the growth experiments performed by Wörsdörfer et al. [180] were repeated using AaLS-QQ, a variant that has the same charge as AaLS-13, but is unable to form capsids. As previously observed, AaLS-13 confers a considerable advantage to cells exposed to HIV-R10 (Figure 56). This advantage only became apparent five hours after inoculation, at which point the cells had already doubled four times. Although in the growth curves recorded by Wörsdörfer et al. cell densities of AaLS-wt and AaLS-neg stagnate around OD_{600} ~0.2, they show a similar five hour delay before a difference between AaLS-wt, AaLS-neg and AaLS-13 becomes visible. The origin of this phenomenon is unknown. Potentially, the detrimental effect of HIV protease may be linked to a cellular response to increased cell density. Nevertheless, AaLS-QQ was not able to rescue *E. coli* from HIV protease toxicity. On the contrary, cells transformed with AaLS-QQ grew slower than those harboring AaLS-wt and AaLS-

neg. When cell growth was assessed in the absence of the toxic protein (see Chapter 5), cells expressing AaLS-13 grew slower than cells expressing AaLS-wt or AaLS-neg. Presumably, the strongly negatively charged proteins interfere with the cellular machinery essential for growth. Likely candidates are positively charged ribosomal proteins.

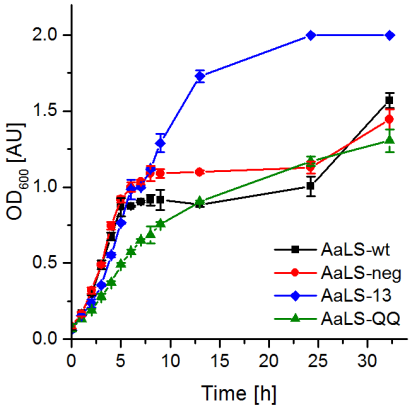


Figure 56. Effect of AaLS variants on cell growth. Overnight cultures of *E. coli* KA13 cells transformed with a plasmid encoding HIV-R10 and a plasmid encoding an AaLS variant were used to inoculate 5 mL liquid LB medium supplemented with 36 µg/ml chloramphenicol, 150 µg/mL ampicillin, 100 µM salicylate and 450 ng/mL tetracycline at 30°C, 230 rpm in sterile plastic tubes. OD_{600} values were not acquired at a path length of 10 mm. Error bars indicate standard deviation from two experiments.

Observation of assembled capsid after purification under low salt conditions (I ≤ 250 mM) in the presence, but not the absence, of positively charged protein guests is key to understanding why AaLS-13 confers an advantage when co-expressed with HIV-R10. These results are consistent with the findings of Wörsdörfer et al. who observed that the presence of GFP36+ enhances the rate of AaLS-13 capsid assembly in vitro [181]. Apparently, interactions with positively charged guest proteins at physiological ionic strength are as effective as high concentrations of salt at reducing the Coulombic repulsion between AaLS-13 subunits and thus enabling capsid assembly.

3.4 Stability of the AaLS-13 capsid and variants

The evolved AaLS-13 capsid contains three cysteine residues: Cys 37 is present already in AaLS-wt and in the crystal structure (Figure 57) [159]. Its thiol group is buried in a hydrophobic patch made up by His 41, Trp 137, Leu 141 and Ile 144. The

other two thiols were introduced during directed evolution of AaLS-neg by R52C and R127C mutations.

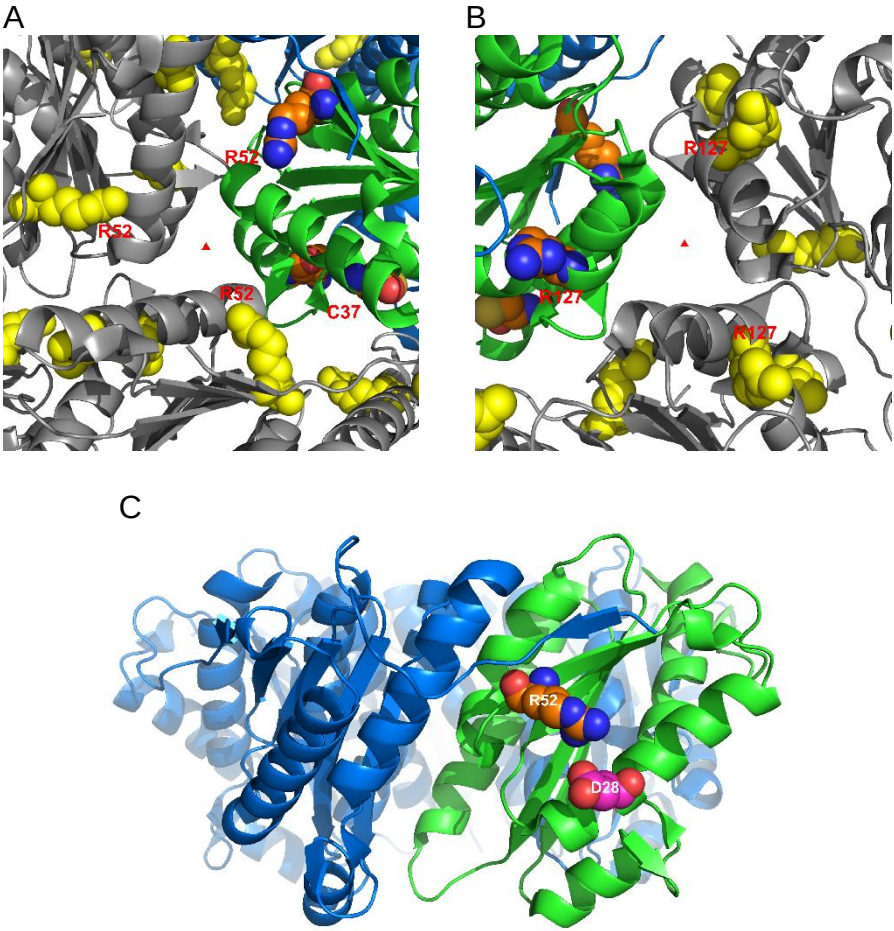


Figure 57. Crystal structure of AaLS-wt (1HQK) [159]. One monomer is highlighted in green, the other four monomers of the same pentamer are shown in blue. Neighboring pentamers are shown in grey. The C_3 symmetry axis is indicated by a triangle. A) View from the capsid exterior along the C_3 symmetry axis. B) View from the capsid interior along the same C_3 symmetry axis. C) Side view of a pentameric subunit.

Arg 127 is situated at the inner, lumenal side of the wild-type capsid. Its guanidinium group forms a salt bridge with Glu 126 (2.8 Å) and is surrounded by Gln 123 (mutated to glutamate in AaLS-neg), His 132 of the same monomer, and Ala 85 of

an adjacent monomer within the same pentamer. Changing this residue to cysteine would not only abolish the salt bridge, but also leave a void. Arg 52 is located on an α-helix approximately half way between the interior and exterior shell surface and points towards the three-fold symmetric axis. It presumably forms a salt bridge with Asp 28 on another α-helix in the same monomer. Asp 28 was mutated to a helix-breaking glycine during the evolution of AaLS-13.

In order to investigate the dynamics of the AaLS-13 capsid, selective labeling of the exterior and interior surface with fluorescent probes was desired. To that end, Cys 52 and Cys 127 were both mutated to either alanine (AaLS-AA), methionine (AaLS-MM), arginine (AaLS-RR) or glutamine (AaLS-QQ). Mutating one surface exposed residue on the interior or exterior to a cysteine would then allow selective labeling of the corresponding face with a thiol-reactive probe. Cys 37 is presumably less reactive than an exposed cysteine and would not be significantly labeled when sub-stoichiometric amounts of reagent are employed. However, changing Cys 52 and Cys 127 to any of the above-mentioned residues renders the capsid kinetically labile under standard conditions (50 mM NaP_i, 200 mM NaCl, 5 mM EDTA, pH 8.0, I = 350 mM, 25 °C). One exception is AaLS-QQ, which does not form capsids at all. Most viral capsids show hysteresis (see section 3.1.2, page 89), i.e. once formed, the capsid does not dissociated into its subunits, even at high dilution [167]. In contrast, re-purification of the isolated capsid fractions of the AaLS-AA, AaLS-MM and AaLS-RR variants affords a re-equilibrated mixture of capsid and pentameric subunits (Figure 58).

The inability of AaLS-RR to form kinetically stable capsids is particularly noteworthy as this variant lies along the evolutionary trajectory from AaLS-neg to AaLS-13. It was therefore studied in more detail. Re-purification of AaLS-RR capsids at different ionic strengths shows that the pentamer-capsid equilibrium, and presumably the kinetic stability of the capsid as well, are dependent on the salt concentration (Figure 59A). Although the equilibrium lies completely on the side of the pentamer at relatively low ionic strength (150 mM, pH 8.0), at an ionic strength of 750 mM and above, virtually the entire protein is present as capsid. Alternatively, at low ionic strength, decreasing the pH can likewise increase capsid stability. Both measures reduce Coulombic repulsion between the subunits: at low pH the proteins become less negatively charged by protonation of the carboxylates; at higher ionic strength the charges are more efficiently screened and the electrostatic interactions thus shorter ranged.

For AaLS-13, dissociation of assembled capsids into subunits was not observed at pH 8.0 at any ionic strength. Dissociation of AaLS-13 capsids can be effectuated by urea with a critical concentration somewhere between 3 and 5 M (Figure 59B).

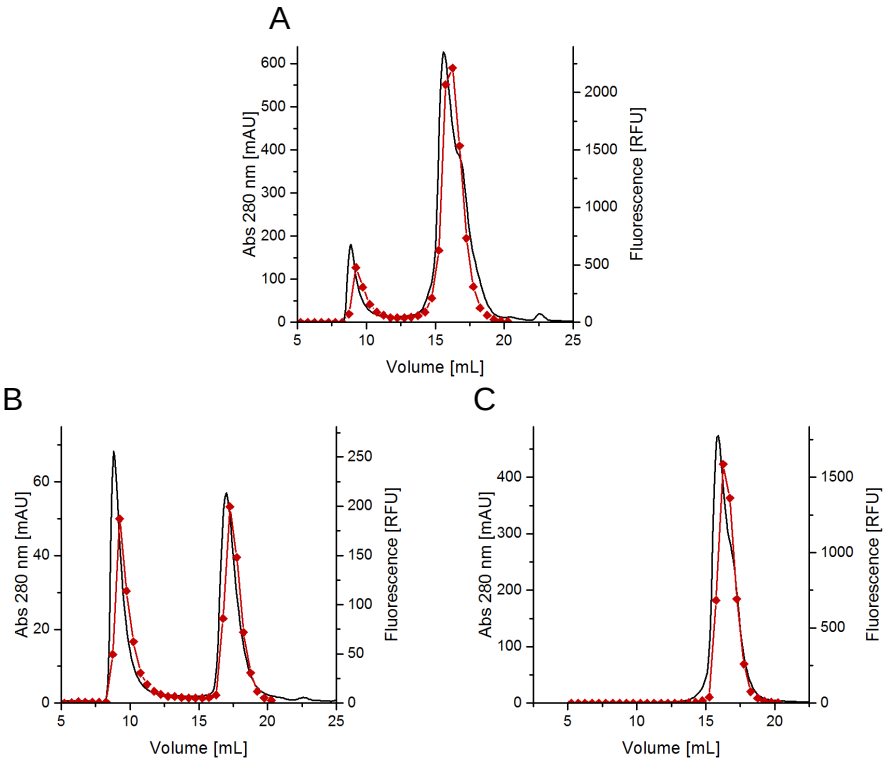


Figure 58. Capsid-pentamer equilibrium of AaLS-AA A85C. A) Capsid (85 µM monomer) was mixed with 8.5 µM GFP36+ and purified by size exclusion chromatography at 50 mM NaP_i, 200 mM NaCl, 5 mM EDTA, pH 8.0. B) Re-purification of the isolated capsid from (A) (fractions from 8.5-11.5 mL) within 1 hour of isolation. C) Re-purification of the isolated pentamers from (A) (fractions from 14.5-18 mL). Black lines indicate on-line absorption at 280 nm, red lines indicate fluorescence of individual fractions (excitation 488 nm, emission 514 nm).

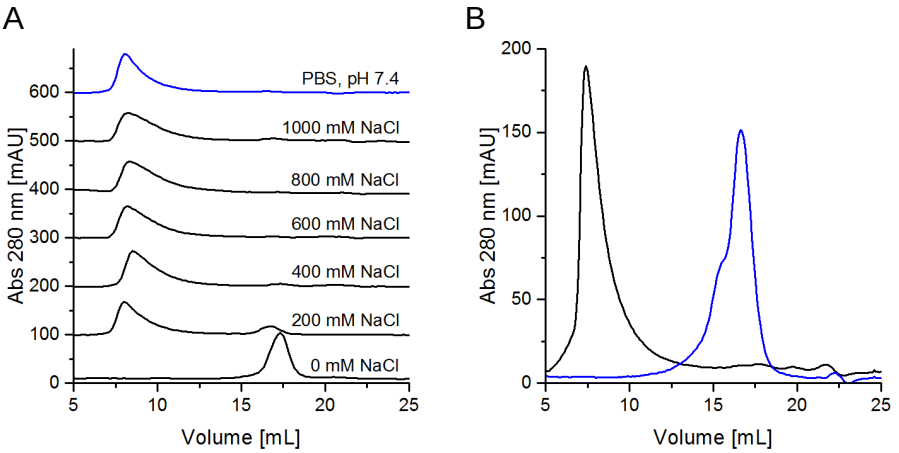


Figure 59. Stability of AaLS capsids. A) The capsid-pentamer ratio for AaLS-RR as a function of ionic strength. AaLS-RR capsids were diluted from a stock solution (2.7 mM, I = 350 mM) to a final concentration of 50 µM into buffer (50 mM NaP_i, 5 mM EDTA, pH 8.0; I = 150 mM) supplemented with NaCl at different concentrations. The samples were subsequently analyzed by size-exclusion chromatography using a Superose 6 10/300 GL equilibrated with the same buffer at room temperature. Additionally, PBS (pH 7.4, I = 160 mM) was tested (blue line). B) Size-exclusion chromatogram of AaLS-13 in 50 mM NaP_i, 200 mM NaCl, 5 mM EDTA, pH 8.0 supplemented with 3 (black) or 5 M (blue) urea. Purified AaLS-13 capsids were diluted to a final monomer concentration of 100 µM and analyzed by size-exclusion chromatography on a Superose 6 10/300 GL column equilibrated with buffer containing the same urea concentration as the samples.

3.5 Quaternary state of AaLS capsids

3.5.1 Size of AaLS capsid variants

The evolution of AaLS-wt to AaLS-neg and AaLS-13 was accompanied by an iterative increase in capsid diameter, evident in transmission electron microscopy images of the individual capsids [177,179]. AaLS-wt forms 60-meric capsids with an outer diameter of 15.4 nm composed of 12 pentamers arranged as a dodecahedron [159]. AaLS-neg capsids have an outer diameter of 28.6 ± 2.6 nm, but in the absence of structural information, the number of subunits it contains is a matter of speculation.

This polymorphism had been previously observed for other lumazine synthase variants. Introducing the tetra-peptide IDEA after Gly 129 in AaLS-wt converted the native capsid into expanded structures with an approximate diameter of 30 nm [205]. This motif had been identified in *Saccharomyces cerevisiae* lumazine synthase and was expected to prevent capsid formation when inserted into AaLS-wt. Capsids of similar size were also observed when lumazine synthase from *Bacillus subtilis* was assembled from subunits in the absence of ligands [206]. A computer

model of a 180-meric, *T=3* icosahedral lumazine synthase capsid has a diameter of ~29 nm and a 240-meric, *T=4* capsid has a diameter of ~32 nm [205]. Simply extrapolating from the surface area of a sphere composed of 60 monomers with d = 15.4 nm to a sphere with d = 28.6 ± 2.6 nm predicts that the AaLS-neg capsid will contain 210 ± 40 monomers. The caveat of such an intellectual exercise is the requirement that a capsid with *T=3* or *T=4* icosahedral geometry must contain hexamer subunits [157]. While some lumazine synthases exist as pentameric assemblies, hexameric structures have never been observed.

In contrast to AaLS-neg, lumazine synthase assemblies the size of AaLS-13 (d = 35.9 ± 2.5 nm) [179] have no precedent. Given its size compared to AaLS-wt, AaLS-13 should be comprised of 330 ± 50 monomers. The intermediate, AaLS-RR, with a diameter of 32 ± 2 nm, would be composed of 260 ± 30 subunits.

3.5.2 Determining AaLS capsid composition by native mass spectrometry

Development of soft electrospray ionization techniques and high resolution mass spectrometers has given rise to the field of structural mass spectrometry (MS) [207,208]. Gentle ionization has made it possible to investigate the composition of biomacromolecular complexes by maintaining the integrity of protein assemblies that are held together by non-covalent interactions upon transfer from solution to the gas phase. Measuring the mass of the intact complex by a mass analyzer of choice allows determination of the composition of the complex. Preservation of the protein quaternary structure is achieved by employing an aqueous but volatile buffer such as ammonium acetate. Briefly, an aerosol of positively charged aqueous droplets, generated by applying an electric field between a capillary and the entrance, is guided into a vacuum chamber. The solvent evaporates under the reduced pressure. In the process, the ammonium ions bearing the positive charge dissociate and their charge is transferred by protonation of the protein complex. Using this strategy and a range of mass analyzing techniques, including time-of-flight [209,210], quadrupole-time-of-flight [211–213], orbitrap [193] and charge-detection mass spectrometry [214,215], the composition of a variety of virus-like particles and bacterial nanocompartments [216] have been determined. The ability to determine the mass of intact 18 MDa capsids from bacteriophage HK97 with an uncertainty of 4 kDa highlights the power of this technology [217].

Purified capsids of the variants AaLS-wt, AaLS-neg, AaLS-RR and AaLS-13 were analyzed with native mass spectrometry by Michiel van de Waterbeemd and Albert Heck at the Universiteit Utrecht, Netherlands. Since AaLS capsids need to be stored at room temperature, even minute amounts of contaminating proteases can significantly degrade the samples. The resulting heterogeneity complicates analysis of

the mass spectra. Unfortunately, addition of protease inhibitor cocktails led to covalent modification of the capsid proteins and interfered with the mass analysis. Since the unstructured C-terminal His6 peptide employed for metal affinity purification is particularly prone to proteolysis, constructs lacking this tag were produced and purified for all four variants. In the case of AaLS-wt and AaLS-neg, most of the endogenous *E. coli* proteins could be precipitated by a heat shock. For AaLS-RR and AaLS-13, a similar enrichment was achieved using ammonium sulfate fractionation. Following sequential ion exchange and size-exclusion chromatography, all capsids were obtained in high purity (Figure 60). With the exception of AaLS-wt, the capsids eluted as symmetric peaks in the order corresponding to their diameter determined by electron microscopy (larger particles first). The *T=1* AaLS-wt capsid corresponds to the species eluting at 74 mL. However, as the only variant lacking substantial negative charge, AaLS-wt capsids tend to aggregate and cause extensive peak fronting. Species corresponding to the fronting peak with a maximum at 66 mL are most likely capsid dimers in equilibrium with higher-order species. Electron microscopy pictures of all AaLS-wt fractions show only a single species with an approximate diameter of 16 nm (Figure 61).

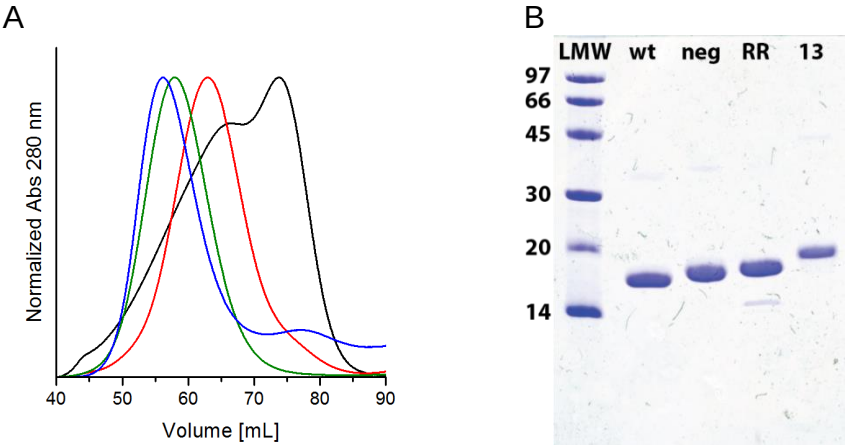


Figure 60. AaLS variants without His6 tag. A) Size-exclusion chromatogram of the final step of capsid purification (HiPrep 16/60 Sephacryl S-400 HR, GE Healthcare). B) SDS-PAGE analysis of purified AaLS variants.

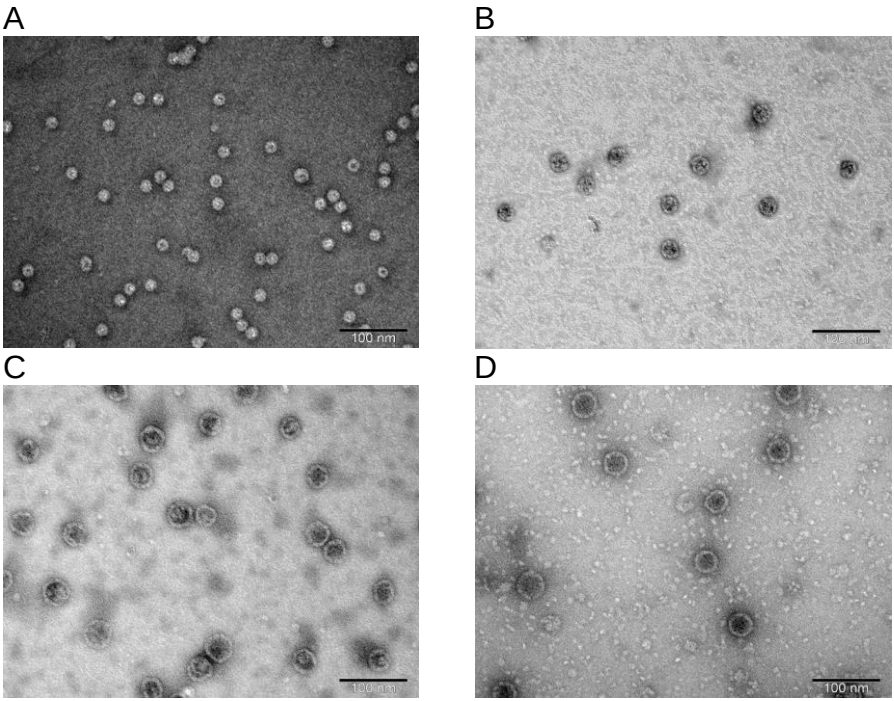


Figure 61. Negative-staining electron microscopy images of AaLS variants A) AaLS-wt, B) AaLS-neg (with guests), C) AaLS-RR and D) AaLS-13. The scale bar is 100 nm. Panel D was recorded by Dr. Eita Sasaki.

Using native mass spectrometry with a quadrupole time of flight mass spectrometer, the masses of AaLS-wt, AaLS-neg and AaLS-RR capsids were determined (Figure 62). Determination of the charge state of each peak through their corresponding m/z position allows measurement of the total mass of the protein complexes. The mass of the AaLS-wt capsid was determined to be 1.056 MDa, corresponding to a 60-mer as expected from the crystal structure. In the case of AaLS-neg, the major charge envelope originates from a species with a mass of 3.017 MDa, corresponding to a 180-meric capsid. For AaLS-RR, electron microscopy images show capsids of at least two distinct sizes (Figure 61D). The mass spectrum of this sample consisted of three charge envelopes, most likely corresponding to a 180-mer (3.013 MDa), a 210-mer (3.518 MDa) and a 240-mer (4.025 MDa). Whereas icosahedrally symmetric capsids with T numbers 3 and 4 can be constructed from 180 subunits and 240 subunits, respectively, the 210-mer is peculiar. It may correspond to a capsid that cannot be described by the Caspar-Klug formal-

ism, possibly a chimera of *T=3* and half *T=4* capsids or an incomplete 240-mer capsid. In case of AaLS-13, the charge states could not be resolved, hinting at a distribution of differently sized capsids. The position of the signals relative to those of AaLS-wt, -neg and -RR suggests that the AaLS-13 capsid contain between 240 and 420 subunits.

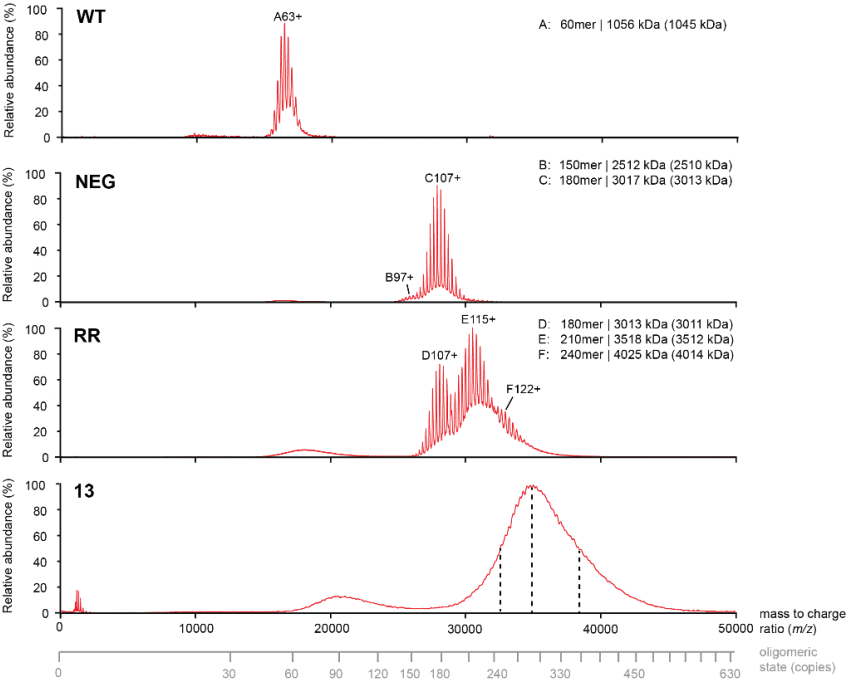


Figure 62. Native-MS spectra of AaLS variants WT, NEG, RR and 13. Deconvolution of the charge states allows to determine the mass of each capsid species (A-F, annotated on the right; expected masses in parentheses). For AaLS-13, charged states could not be resolved. The range of oligomeric states can be estimated by evaluating the position of its signal with respect to those of AaLS-wt, -neg and –RR. Experiments were performed and the figure prepared by Michiel van de Waterbeemd.

Because it appears as a single species in the MS spectrum, AaLS-neg was singled out for detailed studies. Tandem MS experiments in which monomeric subunits dissociate from the complex in the gas phase through collision with inert gas increase the confidence in the charge state assignment, considerably reducing the uncertainty of the total mass and confirming the number of subunits per capsid independent of knowing the monomer mass (Figure 63). Using the same technique, the number of guests per capsid could be determined for GFP36+/AaLS-neg guest-host complexes (Figure 64).

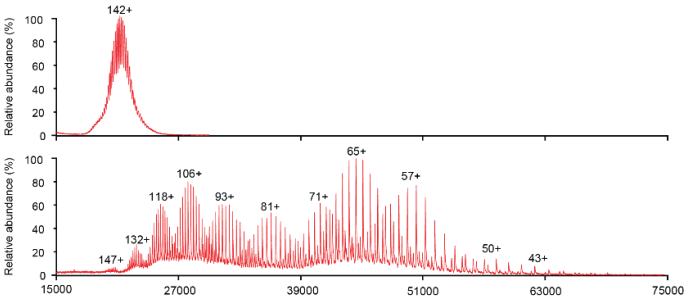
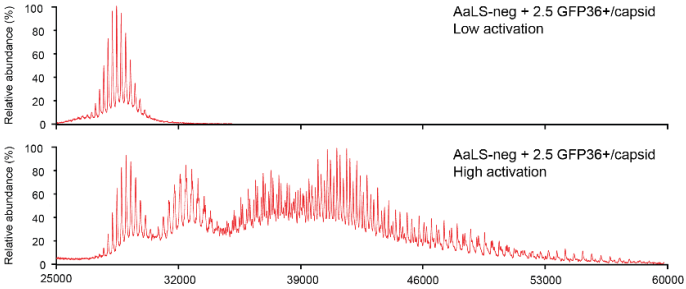


Figure 63. Tandem MS experiments with AaLS-neg. Top: Mass spectrum of intact AaLS-neg capsid. Bottom: spectra from collision-induced dissociation during which pentameric subunits are lost from the capsid. The resulting species bear fewer charges resulting in increased mass/charge ratios. Experiments were performed and figures prepared by Michiel van de Waterbeemd.

A



B

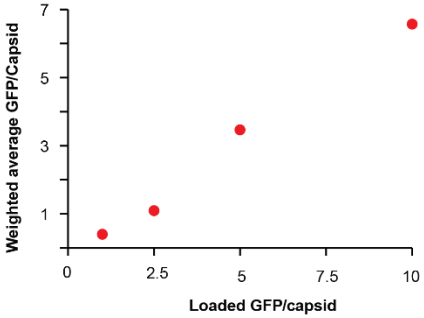


Figure 64. Tandem mass spectra of GFP36+/AaLS-neg complexes. A) Mass spectrum of intact AaLS-neg capsid (top) and fragmentation spectrum (bottom). Analysis of the charge envelopes indicates a mixture of capsids containing either 0, 1, 2 or 3 copies of GFP36+. B) Weighted average of guests per capsid as a function of input ratio. Experiments were performed and figures prepared by Michiel van de Waterbeemd.

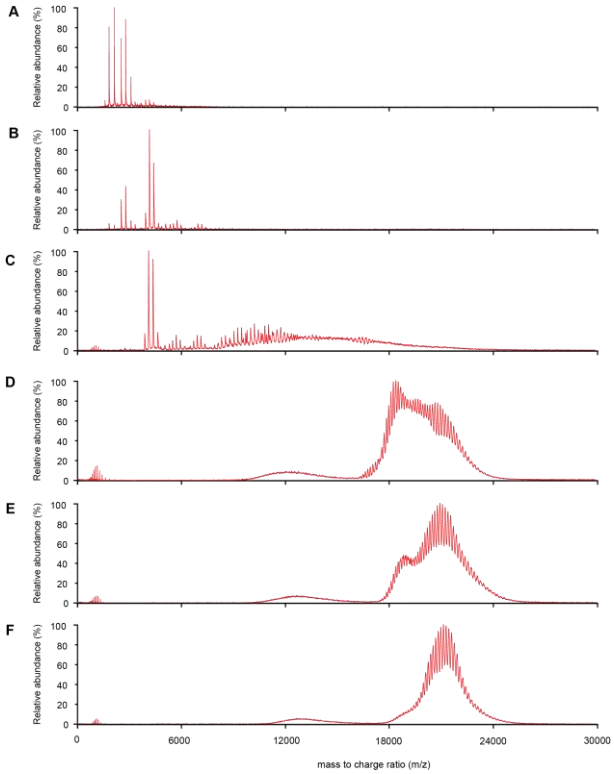


Figure 65. Assembly of AaLS-neg. Monomeric and dimeric species can be detected at low ionic strength and concentrations (A). Increased salt concentration give rise to pentamers (B) and larger intermediates containing multiples of five subunits (C). Late stage intermediates contain approximately 120 and 150 subunits (D and E) which ultimately convert into 180-mers (F). Experiments were performed and figures prepared by Michiel van de Waterbeemd.

Remarkably, several intermediates in the assembly of the AaLS-neg capsid were detected. AaLS-neg was disassembled by changing the buffer to 10 mM ammonium acetate, pH 7.8. After increasing the buffer concentration to 55 mM, the mass spectrum contained signals corresponding to monomers, dimers, pentamers as well as 10, 15, 20, 25, 30, 35 and 40-mers (Figure 65). Addition of trimethylamine, which is a better base than ammonia, reduces the overall protein charge [218] – an effect that was exploited to investigate the very heterogeneous sample containing late-state intermediates in the capsid assembly process (Figure 66). Species with reduced charge appear at higher m/z ratios and the spacing between different charge states increases, facilitating analysis of overlapping envelopes. Thus, signals corre-

sponding to 115, 120, 125, 130, 135, 140, 145, 150, and 155-mers could be assigned. The early as well as the late-state intermediates differ by multiples of five, suggesting that the capsid is assembled by iterative addition of pentamers.

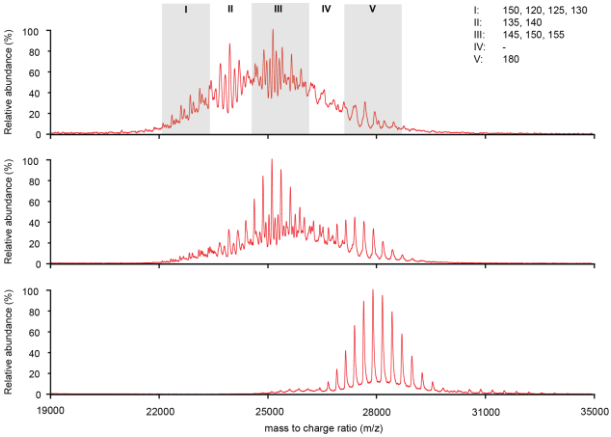


Figure 66. Late intermediates in the assembly of AaLS-neg. Charge reduction allows to resolve the individual species as well as follow their progression to larger species (longer incubation times from top to bottom). Experiments were performed and figures prepared by Michiel van de Waterbeemd.

At a late stage in the assembly process, two species are detected around 19,000 m/z and 21,000 m/z with the former converting into the latter over time. This observation is consistent with an interconversion previously detected by size-exclusion chromatography: at longer incubation times the relative ratio of two species eluting at 65 mL and 74 mL increases, indicating that smaller AaLS-neg species are initially formed during assembly, but these metastable structures are subsequently converted to complete capsids (180-mer) (Figure 67). Different absorbance spectra of both capsid species demonstrate that the smaller species binds flavins, whereas the larger one does not. The negative charges introduced on the lumenal face of the shells presumably weaken the attraction between subunits, rendering the smaller capsid kinetically labile. The reduced Gaussian curvature in expanded capsids would be expected to increase the distance between the negatively charged surfaces of neighboring subunits, decreasing Coulombic repulsion and thus stabilizing the capsid. It would be interesting to test whether AaLS-neg's metamorphosis is slowed down or halted altogether at high ionic strength. In a recent study investigating capsid assembly by resistive-pulse sensing on a nanofluidic device, Zlotnick, Jacobsen and co-workers found that the ratio of smaller *T=3* to larger *T=4* hepatitis B virus-like particles increased with increasing ionic strength [219]. Additionally, the

distribution of large and small AaLS-neg capsids might be influenced by small molecules. Flavins or their precursors might bind to the lumazine active site located at the pentamer-pentamer interface and thus stabilize the *T=1* capsid. For example, the formation of *Bacillus subtilis* lumazine synthase capsids has been shown to be susceptible to selected small molecules [206].

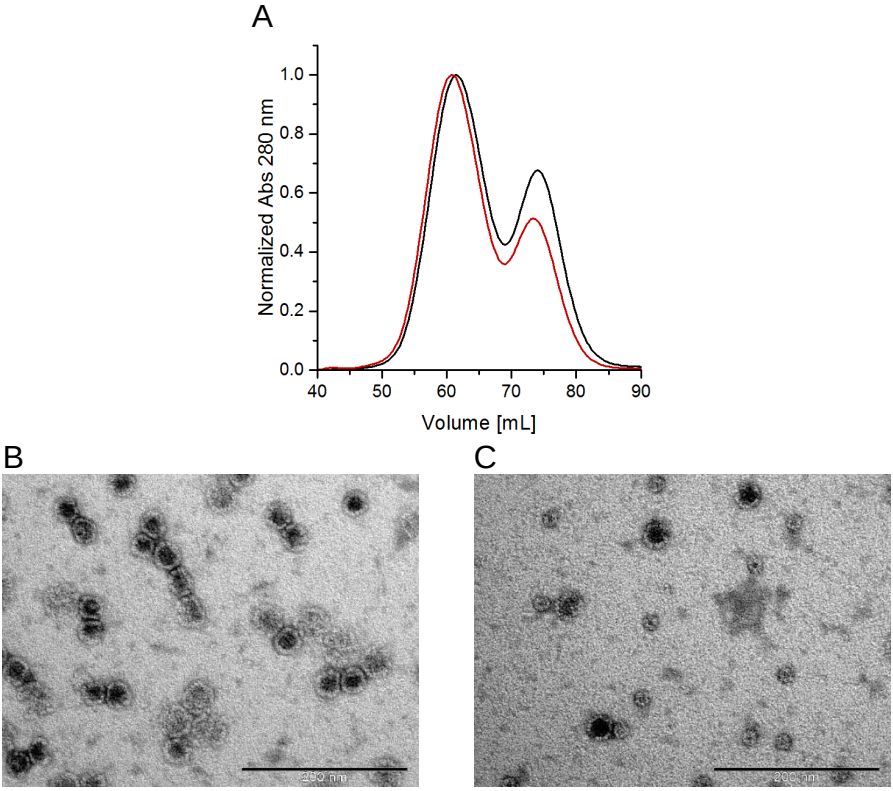


Figure 67. Metamorphosis of AaLS-neg capsids. A) Normalized size-exclusion chromatogram (Sephacryl S-400 HR) of AaLS-neg (without His6 tag) after ion-exchange purification directly (black) and 12 hours later (red). Below: Electron microscopy image of the fraction eluting at 62 mL (B) and 74 mL (C). The smaller capsids have a diameter of 20.3 ± 1.5 nm.

Analysis of AaLS capsids by native mass spectrometry enabled determination of the composition of three of the four investigated variants and shed light onto the assembly of AaLS-neg. Astonishingly, all intermediates in the assembly of AaLS-neg differ by multiples of five subunits, which suggests that pentamers are the universal building block for this capsid. The charge states of AaLS-13 could not be resolved,

presumably due to sample heterogeneity and comparatively low capsid stability under the ionization conditions.

3.6 Summary and outlook

Proteins have evolved over millions, if not billions of years to become extraordinary catalysts, highly selective receptors, or mechanical units with remarkable strength and flexibility. Evolutionary pressures led to optimization of every structural and dynamic property of these proteins, thereby forging the best possible tools for practical applications in the micrometer-sized factory called cell. Scientists are not yet able to reproduce the complexity inherent to living cells, or even the fundamental units from which they are assembled, and have instead begun to modify and adapt natural structures for new ends. Given that nature spent eons to harmonize all features of a protein toward a specific function, it is not surprising that anthropogenic repurposing cannot achieve the same sophistication in the course of a few years. Thus, when enzymes are evolved toward a new, unprecedented function (with remarkable success! [220]), other properties may adjust as well. In the case of P450 enzymes, for instance, mutations frequently decrease coupling efficiency [221]. More generally, the overall protein structure often becomes destabilized [222].

It is thus not surprising that changing lumazine synthase, evolved to catalyze a reaction and encapsulate another enzyme, to a protein container endowed with substantial charge and capable of sequestering complementarily charged guests would result in substantially different physical properties. The initial proposal to employ the evolved capsid as a nano-sized reaction chamber was based on the belief that AaLS-neg and AaLS-13, like viral capsids, form stable, rigid, continuous shells that could provide a considerable diffusion barrier to organic molecules. Given the speed with which proteinaceous cargo is encapsulated and the ease with which it is released, this notion needs revision. At the same time, these unanticipated properties set these supercharged capsids apart and suggest new applications that are not possible employing conventional virus-like particles. The reversibility of the encapsulation process might be leveraged to sequester harmful proteins other than HIV protease, utilizing AaLS-13 as a vault to prevent interaction of the toxin with its cellular target. Using this strategy, toxic, yet useful proteins could be expressed as GFP36+ or deca-arginine fusion proteins in *E. coli*, co-purified by metal affinity chromatography taking advantage of AaLS-13's His6-tag, and subsequently eluted by increasing the ionic strength.

Complementarily, the elution of guest proteins might be exploited for targeted delivery of therapeutic proteins. Both the N- or C-terminus of AaLS can be modified with peptides that specifically bind to the surface of diseased cells [199] and thus

facilitate endocytosis. Capsids loaded with a fusion protein made from GFP36+ and a therapeutic protein of interest, connected by a linker susceptible to changes in pH or redox potential, could then release their cargo upon internalization.

Leakage of GFP36+ from AaLS-13 depends on the strength of Coulombic interactions between guest and host, modulated by ionic strength as well as charges of host and guests. Slow release of GFP(36+)-fusion proteins from AaLS-13 capsids into the blood stream could be explored as a means of achieving sustained drug release. If the linker contained a protease recognition site, temporary exposure to the bulk blood stream could lead to cleavage by an endogenous protease such as blood coagulation factor X_a [223] and release of the encapsulated therapeutic protein. Since the immunogenicity of the proteinaceous capsids is likely a major obstacle for all in vivo applications, PEGylation might be successfully used to attenuate the immune response [224,225].

Although the experiments described in this chapter have advanced our understanding of the AaLS-13 system, many questions remain that might be addressed using the fluorescence spectroscopy techniques introduced. For example, it should be possible to measure the dependence of capsid assembly kinetics on ionic strength and the presence of charged guests using a mixture of AaLS-13 A85C pentamers covalently labeled with two different fluorophores. Monitoring FRET upon addition of salt or charged guests might provide insights into the time course of capsid assembly. Complementarily, titration of assembled capsids composed of pentamers labeled with two different fluorophores with chaotropic agents could provide a means to quantify the thermodynamic stability of the different capsid variants.

3.7 Materials and methods

3.7.1 Molecular biology

General remarks. All AaLS variants used herein contain one additional mutation Q2E in comparison to the sequence deposited in UniProt [226]. This mutation was likely introduced by subcloning AaLS-wt [177]. His6-GFP36+ and His6-TOP36+ contain an additional 9-amino acid C-terminal peptide, SGGSMALER, in comparison to the sequence published by Liu et al. [183]. This peptide was introduced for no apparent reason during construction of plasmid pACYC-GFPpos [181]. When cloning plasmids encoding GFP36+ and TOP36+ without His6-tag (see below) this peptide was removed as well, because it is easily proteolyzed and thus introduces unwanted heterogeneity.

pACYC_His6-TOP36p. Iterative overlap extension PCR was using pACYC_fw and pACYC_rv as flanking primers and Top1_fw, Top1_rv, Top2_fw, Top2_rv, Top3_fw and Top3_rv (Table 4) for introduction of four site-directed mutations based on pACYC_His6-GFP(36+) [227]. The resulting gene was digested with *Bam*HI and *Xho*I and subsequently ligated into digested and dephosphorylated pACYC_His6-GFP(36+) to afford pACYC_His6-TOP36p.

pACYC_GFP36p and pACYC_TOP36p. Genes encoding for GFP36+ and TOP36+ were amplified from plasmids pACYC_His6-GFP36p and pACYC_His6-TOP36p, respectively, using primers NdeI_GFP36p_fw and GFP36p_rv. The gene products were digested with *Nde*I and *Spe*I and subsequently ligated into digested and dephosphorylated pACYC_His6-GFP36p to afford pACYC_GFP36p or pACYC_TOP36p.

pACTET_TOP36p_GFP36p. The gene encoding TOP36+ was amplified from pACYC_TOP36p using primers NdeI_GFP36p_fw and RBS_rv. The product is digested with *Nde*I and *Xho*I and subsequently ligated into accordingly restricted pACYC_tetHIV-R10 [179] yielding pACTET_TOP36p. The gene encoding GFP36+ is amplified from pACYC_His6-GFP36p [227] with primers RBS_fw and GFP36p_rv. The product is digested with *Xho*I and *Spe*I and ligated into accordingly restricted pACTET_TOP36p to afford pACTET_TOP36p_GFP36p (Figure 68).

pMG211_AaLS13_A85C was constructed from pMG211_AaLS13 [179] by overlap extension PCR using AaLS_fw and AaLS_rv as flanking primers and A85C_fw and A85C_rv to introduce mutation A85C. The gene product was digested with *Nde*I and *Xho*I and subsequently ligated into digested and dephosphorylated pMG211 [228] to afford pMG211_AaLS13_A85C.

pMG211_AaLS13_C52A_C127A was a generous gift from Dr. Eita Sasaki.

Table 4. Oligonucleotides used for molecular biology.

Name	Sequence
A85C_fw	ATTGGCGTTCTCATCGAAGGGTGCGAGCCACATTTCG
A85C_rv	CCCTTCGATGAGAACGCCAATTGC
AaLS13_fw	AGATATACATATGGAAATCTACGAAGG
AaLS_noHis_rv	GGTGGTGCTCGAGTCATTATTATCGGAGAGACTTGAATAAGTTTGC
AaLS13_rv	TGGTGCTCGAGTCGGAGAGACTTGAATAAGTTTGC
C127M_fw	CGAATTGGAAGAGGCTATCGAGATGGCCGGCACAGAACACGGCAAC
C127Q_fw	CGAATTGGAAGAGGCTATCGAGCAGGCCGGCACAGAACACGGCAAC
C127R_fw	CGAATTGGAAGAGGCTATCGAGCGCGCCGGCACAGAACACGGCAAC
C127R_rv	CTCGATAGCCTCTTCCAATTCG
C52M_fw	CGTGAAGAAGACATTACTCTGGTTATGGTTCCAGGCTCATGGGAAATACC
C52Q_fw	CGTGAAGAAGACATTACTCTGGTTCAGGTTCCAGGCTCATGGGAAATACC
C52R_fw	CGTGAAGAAGACATTACTCTGGTTCGCGTTCCAGGCTCATGGGAAATACC
C52R_rv	AACCAGAGTAATGTCTTCTTCACG
GFP36p_rv	TCAGCTGACTAGTCACTTGTAGCGTTCGTCGCGTCC
NdeI_GFP36p_fw	GATATACATATGGCTAGCAAAGGTGAACG
pACYC_fw	ACTTTTATCTAATCTCGACATCATTAATTCC
pACYC_rv	AGCGGTGGCAGCAGCCAACTCAGC
R108C_fw	AAACCTATCAGCTTCGGTGATATTAC
R108C_rv	GTAATATCACCGAAGCTGATAGGTTTACATAGTTCTAATGAAAGGTTCGCG
RBS_fw	GCTACAAGTGATAACCTCGAGCTAGAAATAATTTTGTTTAACTTTAAGAAGG
RBS_rv	TCTAGCTCGAGGTTATCACTTGTAGCGTTCGTCGCGTCC
T7term	TGCTAGTTATTGCTCAGCGG
Top1_fw	TATGGCGTTCAGTGCTTCGCGCGTTACCCTAAACACATG
Top1_rv	GAAGCACTGAACGCCATAGCCCAGGGTGGTCACTAAGG
Top2_fw	CGTTACCCTAAACACATGCGCCGTCACGATTTTTTCAAATCAGC
Top2_rv	CATGTGTTTAGGGTAACGCGC
Top3_fw	CGCAACCACTACCTGAGCTATCGTTCTAAACTGAGCAAA
Top3_rv	GCTCAGGTAGTGGTTGCG

pMG211_AaLS13_C52R_C127R was constructed from pMG211_AaLS13 by overlap extension PCR using AaLS_fw and T7term as flanking primers and C52R_fw, C52R_rv, C127R_fw and C127R_rv to introduce mutations C52R and C127R. The gene product was digested with *Nde*I and *Xho*I and subsequently ligated into digested and dephosphorylated pMG211 to afford pMG211_AaLS13_C52R_C127R. Plasmids ***pMG211_AaLS13_C52M_C127M*** and ***pMG211_AaLS13_C52Q_C127Q*** were prepared analogously using primer pairs C52M_fw/C52R_rv, C127M_rv/C127R_rv, C52Q_fw/C52R_rv, and C127Q_rv/C127R_rv respectively.

pMG211_AaLS13_C52R_A85C_C127R was constructed analogously using pMG211_AaLS13_A85C as template.

pMG211_AaLS13_C52R_R108C_C127R was constructed from pMG211_AaLS13_C52R_C127R using AaLS_fw and AaLS_rv as flanking primers and R108C_fw and R108C_rv to introduce mutation R108C. The gene product was digested with *Nde*I and *Xho*I and subsequently ligated into digested and dephosphorylated pMG211 to afford pMG211_AaLS13_C52R_C127R_R108C.

pMG211_AaLS*_noHis. Plasmids encoding AaLS variants wt, neg, RR and 13 without His6-tag were constructed by amplifying the respective genes encoding the His6-tagged variant [177,179] with primers AaLS_fw and AaLS_noHis_rv. The product was digested with *Nde*I and *Xho*I and subsequently ligated into digested and dephosphorylated pMG211 to afford pMG211_AaLS*_noHis.

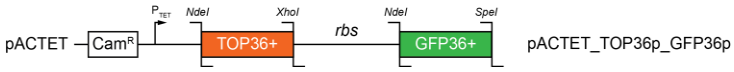


Figure 68. Structure of the pACTET_TOP36p_GFP36p plasmid used for co-expression of TOP36+ and GFP36+ under control of the tetracycline repressor.

3.7.2 Protein production and purification

GFP36+ and TOP36+. A dense overnight culture of BL21 gold DE3 carrying either plasmid pACYC_His6-GFP36p, pACYC_His6-TOP36p, pACYC_GFP36p or pACYC_TOP36p was used to inoculate 400 mL of LB medium supplemented with 36 µg/mL chloramphenicol and incubated at 37 °C and 230 rpm until OD_{600} reached 0.6-0.7. Following addition of 100 µM IPTG, the temperature was reduced to 25 °C and incubation continued for 20 h. Cultures were centrifuged at 4 °C and 4,000 g for 20 min, and the resulting cell pellet stored at -20 °C. The cell pellet was thawed and suspended in lysis buffer (50 mM NaP_i, 2 M NaCl, pH 7.4) supplemented with 1 mg/mL lysozyme, protease inhibitor cocktail (cOmplete protease inhibitor cocktail tablets, Roche, Switzerland) and a spatula tip of both DNase I and RNase A. Following incubation at room temperature for 1 h, cells were lysed by sonication and

the insoluble fraction removed by centrifugation at 4 °C and 14,000 g for 30 min. In case of His6-tagged proteins, the clear lysate was loaded on 4 mL of Ni-NTA agarose resin equilibrated with lysis buffer. Following washing with lysis buffer and lysis buffer containing 20 mM imidazole, the protein was eluted with lysis buffer containing 250 mM imidazole. Protein-containing fractions were concentrated and the buffer changed to 50 mM TRIS-HCl, 500 mM NaCl, pH 7.4 using a centrifugal ultrafiltration unit (10 kDa MWCO, Amicon Ultra, Merck Millipore). In case of GFP36+ and TOP36+ without His6-tag, the cleared lysate was incubated for 5 min at 70°C, cleared again by centrifugation prior to buffer exchange and concentration. The concentrated protein was further purified by ion exchange chromatography using a Mono S HR 10/10 column (GE Healthcare Life Sciences, UK) that had been equilibrated with 50 mM TRIS-HCl, 500 mM NaCl, pH 7.4 at 4 °C using a linear gradient to 50 mM TRIS-HCl, 1.5 M NaCl, pH 7.4. Protein-containing fractions were combined, concentrated, exchanged to 50 mM NaP_i, 600 mM NaCl, 5 mM EDTA, pH 8.0 and stored at 4 °C.

AaLS-13 and variants thereof. A dense overnight culture of BL21 gold (DE3) carrying plasmid pMG211_AaLS13 or a variant thereof was used to inoculate 400 mL of LB-medium supplemented with 150 µg/mL ampicillin and incubated at 37 °C and 230 rpm until OD_{600} reached 0.6-0.7. Following addition of 100 µM IPTG, the temperature was reduced to 30 °C and incubation continued for 20 h. Cultures were centrifuged at 4 °C and 4,000 g for 20 min and the resulting cell pellet stored at -20 °C. The cell pellet was thawed and suspended in lysis buffer (50 mM NaP_i, 300 mM NaCl, pH 8.0) supplemented with 1 mg/mL lysozyme, protease inhibitor cocktail (cOmplete protease inhibitor cocktail tablets, Roche, Switzerland) and a spatula tip of both DNase I and RNase A. Following incubation at room temperature for 1 h, cells were lysed by sonication and the insoluble fraction removed by centrifugation at 25 °C and 14,000 g for 30 min. The cleared lysate was loaded on 4 mL of Ni-NTA agarose resin equilibrated with lysis buffer. Following washing with lysis buffer containing 20 mM or 40 mM imidazole, the protein was eluted with lysis buffer containing 500 mM imidazole. Immediately after elution, 5 mM EDTA was added to the protein solution.

In order to obtain assembled capsids, buffer was changed to 50 mM NaP_i, 600 mM NaCl, 5 mM EDTA, pH 8.0 using a centrifugal ultrafiltration unit (30 kDa MWCO, Amicon Ultra, Merck Millipore). The protein was incubated at a concentration of 1-2 mM for 5-7 days at room temperature before separating capsids from fragments (mostly pentamers) by size-exclusion chromatography using a HiPrep 16/60 Sephacryl S-400 column (GE Healthcare Life Sciences, UK). To obtain the capsid fragments, buffer was changed to 50 mM NaP_i, 100 mM NaCl, 5 mM EDTA and the quaternary states were separated without delay. In the case of AaLS-13 variants

containing an additional cysteine residue intended for chemical modification, 5 mM TCEP was added. Purified protein was stored in either 50 mM NaP_i, 100 mM NaCl, 5 mM EDTA, pH 8.0 (fragments) or 50 mM NaP_i, 200 mM NaCl, 5 mM EDTA, pH 8.0 (assembled capsids) at room temperature.

AaLS-13 and variants thereof without His6-tag. *E. coli* BL21 DE3 gold was transformed with plasmids encoding the AaLS-variants without His-tag. Gene transcription by T7 polymerase is controlled with the *lac* repressor. 400 mL sterile LB containing 150 µg/mL ampicillin was inoculated with 400 µL of a dense overnight culture and incubated at 37°C, 230 rpm until OD_{600} reached 0.6. At this point, protein production was induced by addition of 100 µM IPTG and incubation continued at 30°C for 20 hours. Cells were harvested by centrifugation, washed with PBS and stored at -20°C.

Following thawing, cell pellets were suspended in 15 mL lysis buffer (50 mM NaP_i, 300 mM NaCl, pH 8.0) supplemented with 1 mg/mL lysozyme, 1 mg/mL polymixin, EDTA-free protease inhibitor (cOmplete protease inhibitor cocktail tablets, Roche, Switzerland) and a spatula tip of each DNase I and RNase A. After incubating at room temperature for one hour, cells were disrupted by sonication and cleared by centrifugation for 30 min at 25,000 g, room temperature. The supernatants containing AaLS-wt or AaLS-neg were incubated at 85°C for 20 min, while ¾ volumes of saturated aqueous ammonium sulfate were added to the supernatants containing AaLS-RR or AaLS-13 at room temperature. The resulting suspensions were again cleared by centrifugation. SDS-PAGE analysis revealed that for all four variants the majority of the protein remained in the soluble fraction. For AaLS-RR and AaLS-13, the buffer was exchanged to 50 mM NaP_i, 200 mM NaCl, pH 8.0 by repeated centrifugal ultrafiltration (30 kDa MWCO, Amicon Ultra, Merck Millipore) and dilution. Crude lysates were filtered to remove large particles, applied to a MonoQ HR 10/30 (GE Healthcare) column equilibrated with 50 mM NaP_i, 200 mM NaCl, pH 8.0 at room temperature and eluted by a linear gradient increasing the NaCl concentration to 1 M. Fractions containing AaLS and displaying an appropriate UV-spectrum (A260/A280 < 0.7) were combined, buffer exchanged to 50 mM TRIS-HCl, 1M NaCl, pH 8.0, and concentrated until the monomer concentration exceeded 1 mM. Following incubation over 7 days at room temperature, the protein samples were purified by size-exclusion chromatography using a HiPrep 16/60 Sephacryl S-400 HR column (GE Healthcare) equilibrated with 50 mM TRIS-HCl, 1M NaCl, pH 8.0 at room temperature. Fractions in the middle of the main peak and in case of AaLS-wt eluting after 75 mL were combined and concentrated. In case of AaLS-13, 3 mM DTT was added to all buffers to prevent disulfide-bond formation.

Covalent labeling of AaLS-13 variants. Reagents for covalent labeling of AaLS-13 variants with fluorophores were selected such that modification would occur

selectively at cysteine residues introduced by site-directed mutagenesis. In order to promote labeling of the strongly negatively charged protein, positively charged ATTO-495-maleimide and neutral ATTO-565-maleimide (ATTO-Tec, Germany) were chosen. Between 0.2-0.3 eq. of reactive dye were added to a 50-300 µM solution of AaLS-13 A85C or AaLS-13 C52R/A85C/C127R in 50 mM NaP_i, 200 mM NaCl, 5 mM EDTA, pH 8.0, and the resulting mixture incubated at room temperature in the dark for one hour before quenching the reaction by addition of 1 mM β-mercaptoethanol. Small molecules were removed using a PD-10 column (GE Healthcare Life Sciences, UK). Quarternary states were subsequently separated by size-exclusion chromatography using a HiPrep 16/60 Sephacryl S-400 column (GE Healthcare Life Sciences, UK). The extent of labeling was calculated from UV-VIS spectra of the labeled proteins using ε_{565nm}(ATTO-565) = 120,000 M^{-1} cm^{-1}, ε_{280nm}(ATTO-565) = 19,200 M^{-1} cm^{-1}, ε_{495nm}(ATTO-495) = 80,000 M^{-1} cm^{-1}, ε_{280nm}(ATTO-495) = 31,200 M^{-1} cm^{-1} and ε_{280nm}(AaLS) = 13,980 M^{-1} cm^{-1}. Typically 10-16 % of the monomers were labeled.

Labeling AaLS-13 C52R/R108C/C127R bearing an exposed thiol on the exterior surface with ATTO-565-maleimide under identical conditions resulted in a significantly lower labeling ratio (4 %). In order to improve labeling, the reaction was carried out under modified conditions in 50 mM NaP_i, 1 M NaCl, 5 mM EDTA, pH 8.0, using 2.0 eq. of reagent for an extended incubation time of 12 h, yielding 27-55 % labeled monomers (AaLS-RR*^{OUT}).

3.7.3 Concentration determination

Protein concentrations were determined from UV-VIS absorption spectra between 250-800 nm recorded on a Lambda 20 or Lambda 35 spectrophotometer (Perkin Elmer, USA). For samples containing capsids, the contribution of particle-induced light scattering was taken into account by fitting eq. (1) to regions of the spectra devoid of absorption bands (350-600 nm for empty, unlabeled capsids, 550-800 nm for capsid-cargo complexes and 600-800 nm for chemically labeled capsids) using a|e - UV-Vis-IR Spectral Software 2.0 (FluorTools, www.fluortools.com).

$$A_{scatter}(\lambda) = A_0 + log\frac{1}{1 - c \cdot \lambda^4} \tag{23}$$

By virtue of their distinctive absorption spectra, the concentrations of the individual components in mixtures of AaLS-13, GFP36+ and TOP36+ can be determined by solving a system of linear equations accounting for the individual contributions of each component to the total absorption at three different wavelengths. The extinction coefficient for AaLS-13 was calculated with ε_{280nm}(AaLS-13) = 13,980 M^{-1} cm^{-1}, using ProtPram (http://expasy.org/ tools/protparam.html). Extinction coefficients for GFP36+ are ε_{280nm} = 17,420 M^{-1} cm^{-1}, ε_{488nm} = 36,600 M^{-1} cm^{-1} and ε_{513nm} =

6,010 $M^{-1}cm^{-1}$ [227]. The extinction coefficients for TOP36+ were determined to be ε_{280nm} = 20,610 M^{-1} cm^{-1}, ε_{488nm} = 28,610 M^{-1} cm^{-1} and ε_{513nm} = 69,800 M^{-1} cm^{-1} as previously described [227].

3.7.4 Fluorescence measurements

Fluorescence spectra were recorded on a QuantaMaster 7 (Photon Technology International, USA) at 25 °C.

Quantum yields. Quantum yields were measured at 25 °C on a Fluorolog 3 spectrofluorimeter (Horiba, Japan) using excitation and emission correction. Quantum yield values for GFP36+ and TOP36+ were determined by the reference method with fluorescein (Fluka, Switzerland) in 0.1 M NaOH and rhodamine 6G (Fluka, Switzerland) in EtOH as reference standards with reported quantum yields of 0.91 ± 0.05 and 0.94, respectively [229]. Diluted solutions of the fluorescent proteins were prepared in either 50 mM NaP_i, pH 8.0 or 50 mM TRIS-H_2SO_4, pH 8.0 so that the optical density at the excitation wavelength was below 0.05 AU. No difference in quantum yield was observed between these two buffers. Use of NaCl was avoided as halides are known to quench the fluorescence of YFP variants [188]. Different dilutions of fluorescein and GFP36+ were excited between 410-450 nm and the area of the emission peak (460-660 nm) was plotted as a function of the optical density at the excitation wavelength. Similar plots were prepared for TOP36+ (excitation 420-460 nm; emission 470-670 nm) and rhodamine 6G (excitation 450-490 nm; emission 500-700 nm). Quantum yields of GFP36+ and TOP36+ were calculated by relating the respective slopes to fluorescein. Comparison between fluorescein and rhodamine 6G was used to validate the procedure.

Estimation of FRET efficiencies. Since the absorption and emission spectra as well as the quantum yields of GFP36+ and TOP36+ are similar to their uncharged precursors [187], their Förster distance can be approximated to be 5.6 nm as determined for GFP-YFP [189].

Placing n points on a sphere such that their mutual distances R are maximized is referred to as the Tammes problem, and the solutions for n = 2-130 were retrieved from Tables of Spherical Codes, published electronically at NeilSloane.com/packings [230]. For n > 130, the angular separation φ was approximated with Tóth's formula, eq. (24).

$$\varphi \leq \sqrt{4 - csc^2\left[\frac{\pi\, n}{6(n-2)}\right]} \tag{24}$$

The radius of the sphere on which the chromophores are placed when encapsulated in AaLS-13 was estimated to be 13 nm by subtracting the thickness of the wall (5.0 nm) and the radius of GFP36+ (1.5 nm) from the outer radius of a filled

capsid (19.5 nm).[181] The loading capacity of AaLS-13 is about 0.2 GFP36+ per monomer. One GFP36+ per pentamer corresponds to 12 guests per 60mer (T=1), 36 guests per 120mer (T=3) and so on, resulting in higher fluorophore density and thus FRET efficiency if a larger triangular number is assumed for a capsid of defined diameter.

FRET-efficiency, E_0, between one donor and acceptor is calculated by eq. (25) and FRET-efficiency, E_n, for one donor and n acceptors by eq. (26) [190].

$$E_0 = \frac{1}{1 + \left(\frac{R}{R_0}\right)^6} \tag{25}$$

$$E_n = \frac{nE_0}{1 + (n-1)E_0} \tag{26}$$

The fluorescence emission was decomposed into contributions from GFP36+ (I_D) and TOP36+ (I_A), using a|e - UV-Vis-IR Spectral Software 2.0 (FluorTools, www.fluortools.com). Crosstalk (direct excitation of TOP36+) was determined from the emission spectrum of a GFP36+/TOP36+ mixture in the absence of capsid and the correction included in the equation for ratiometric FRET [231]. Since donor and acceptor emission are marginally separated and quantum yields not significantly different, no corresponding corrections were employed to calculate the FRET efficiency with eq. (27).

$$E = \frac{I_A \cdot \left(1 - \frac{I_A^{free}}{I_D^{free} + I_A^{free}}\right)}{I_D + I_A \cdot \left(1 - \frac{I_A^{free}}{I_D^{free} + I_A^{free}}\right)} \tag{27}$$

pH and ionic strength dependence. Empty AaLS-13 capsids were mixed with a 1:4 mixture of GFP36+/TOP36+ at an ionic strength of 350 mM, pH 8.0, 25 °C and purified by size-exclusion chromatography under the same conditions. Buffers were prepared using MES (pH 6-6.5), HEPES (pH 7-8), CHES/HEPES (pH 8.5), CHES (8.5-9.5), CHES/CAPS (pH 10) or CAPS (pH 10.5-11) and pH was adjusted by addition of NaOH. Concentrated capsid-guest complexes or a mixture of guests were diluted into a solution containing 25 mM buffer and 200 mM NaCl.

A solution containing purified capsid-cargo complex in 50 mM NaP_i, 5 mM EDTA, pH 8.0 was titrated with 50 mM NaP_i, 4 M NaCl, 5 mM EDTA, pH 8.0 containing an identical concentration of the same complex in order to keep concentrations of

proteinaceous solutes constants when determining the effect of ionic strength on the FRET efficiency between encapsulated GFP36+ and TOP36+.

3.7.5 Kinetic measurements

Guest exchange between capsids. Buffers of different ionic strength were prepared by mixing 50 mM NaP_i, 5 mM EDTA, pH 8.0 and 50 mM NaP_i, 1 M NaCl, 5 mM EDTA, pH 8.0 in appropriate ratios. After mixing AaLS-13 containing GFP36+ and AaLS-13 containing Top36+ in a 1:4 ratio at 25 °C, the time course of fluorescence emission was followed at 508 nm or 525 nm upon irradiation at 400 nm. The rate constant for guest exchange k was determined by fitting the trace at 525 nm to eq. (28) in which the linear term accounts for time-dependent photobleaching of the chromophores.

$$F(t) = F_0 + \Delta\mathrm{F} \cdot e^{-kt} + G \cdot t \tag{28}$$

Stopped flow measurements. Guest encapsulation in AaLS-13 variants was observed by following the change in fluorescence emission upon mixing supercharged protein guests with empty capsids in a SX18-MV stopped-flow spectrometer (Applied Photophysics, UK) equilibrated with a water bath at 25 °C. For each transient, 1,000 data points were recorded using a logarithmic time base over 10 seconds, and 8-10 transients were averaged for each experiment. Due to the availability of less material, only 3 transients were averaged for experiments involving chemically labeled AaLS-13 variants.

For encapsulation experiments with mixtures of GFP36+ and TOP36+, the mixing cell was irradiated at 400 nm and fluorescence emission detected at a 90° angle. GFP36+ emission was selectively detected using a band pass filter with center wavelength at 500 nm and a full width at half maximum (FWHM) of 10 nm (Laser Components GmbH, Germany); a long pass filter with an edge at 515 nm (Applied Photophysics, UK) was used for TOP36+ emission. For encapsulating TOP36+ in AaLS-13 variants labeled with ATTO-565, irradiation occurred at 450 nm and TOP36+ emission was selectively detected using a band pass filter with a center wavelength at 525 nm and FWHM of 25 nm (Edmund Optics Inc., USA) while a long pass filter with an edge at 570 nm (Applied Photophysics, UK) was used for ATTO-565 emission. Light scattering was detected either by absorption at 400 nm or irradiation between 600-650 nm combined with a photomultiplier positioned at a 90° angle. Both methods were used in conjunction to confirm that signal changes arose from light scattering by colloidal particles and not changes in the light-transmissive or fluorescent properties of the chromophores. The slit widths of the monochromator used to select the wavelength for irradiation and the photomultiplier voltage were adjusted for each experiment and signal channel to optimize the signal-to-

noise ratio. Consequently, the amplitudes of transients cannot be compared between experiments, unless explicitly stated.

At low protein concentrations for which intermediate aggregation was not observed, second-order kinetics were assumed for the association between guest proteins and binding sites within the capsid (a binding site spans approximately 5 capsid monomers). If the starting concentrations of both reactants are set equal, the time dependence of the fluorescence emission can be described by eq. (29).

$$F(t) = F_0 + \Delta F \cdot \frac{{c_0}^2 k_1 t}{1 + c_0 k_1 t} \tag{29}$$

Even if the initial concentrations of the two components are not exactly identical, this equation serves as a good approximation in which c_0 is calculated as the average of guest and binding site concentration [232]. For different starting concentrations, c_0, eq. 29 is fitted to the progress curve of TOP36+ emission to determine $c_0 k_1$. Linear regression of $c_0 k_1$ against c_0, yields the bimolecular rate constant for association k_1.

3.7.6 Calculation of the electrostatic surface potential

The crystal structure of wild-type AaLS, PDB code 1HQK [159], was relaxed under symmetry constraints [233] using Rosetta 3.4 [234] before introducing mutations for AaLS-13 by *fixbb*, followed by another round of relaxation. The resulting structure was used as input to calculate the surface electric potential at I = 150 mM, pH 8.0 using the Adaptive Poisson-Boltzmann Solver [53].

3.7.7 Electron microscopy

Samples were diluted to 4 µM AaLS in standard buffer. Carbon coated copper grids (400 mesh, Ted Pella Inc., 01814-F) were treated with glow discharge prior to placing 4 µL sample on the grid and incubating for 1 min. The grids were washed three times in standard buffer, then one time in staining solution (2 % uranyl acetate) prior to incubating 20 seconds in staining solution. The grids were air dried before analysis using a FEI Morgagni 268 transmission electron microscope, operated at 100 kV.

3.7.8 Guest-host co-expression

3.7.8.1 GFP36+/TOP36+ and AaLS-variants

Plasmids pACTET_TOP36p_GFP36p and pMG211_AaLS* were co-transformed into *E. coli* strain KA13 [203]. Colonies of KA13 cells co-transformed with pMG211_AaLS variants and pACTET_Top36p_GFP36p were used to inoculate 5 mL LB (25 µg/mL chloramphenicol (Cam), 100 µg/mL ampicillin (Amp)) and incubated

at 30 °C. After 11 hours, OD_{600} values were determined by diluting the cell culture 10-fold into PBS. OD_{600} values of the cultures ranged between 0.7 and 2.1. The cultures were diluted into LB (25 µg/mL Cam, 100 µg/mL Amp) to reach a final OD_{600} of 0.2. Each 1.1 mL culture was added to one well of a 96-deep-well plate and the solution incubated at 37 °C, 1,000 rpm for 45 min followed by 45 min at 30°C. OD_{600} and fluorescence measurements were performed in a plate reader after transferring 50 µL of cultures to clear (absorbance) and black (fluorescence) plates that contained already 125 µL PBS per well to yield a final volume of 175 µL. After the initial measurement (t=0), 1 mL culture was left and 10 µL tetracycline was added from differently concentrated stocks in EtOH yielding final concentrations between 0 and 5,000 ng/mL. Measurements were made 1, 3 and 4.5 hours after addition of tetracycline. After 5 hours, production of AaLS variants was induced by addition of 100 µM IPTG and time points were analyzed at 6.5, 11.5 and 21 hours. After 3 hours, sufficient GFP36+ and Top36+ had been produced to optimize the fluorescence measurements. Excitation spectra were recorded upon excitation at different wavelengths between 400-470 nm. At low wavelengths, the cellular background is too high, whereas at high wavelengths GFP36+ is not selectively excited and the emission maximum is already at 525 nm (Top36+ emission). Excitation at 420 nm was found to be optimal for detecting FRET between GFP36+ and TOP36+ within *E. coli* cell cultures. To visualize the composition of the proteome by SDS-PAGE, 300 µL cell culture were pelleted at 14,000 g and the pellet re-suspended in 60 µL BugBuster reagent (Novagen, Merck, Germany), supplemented with 1 µL benzonase (Merck, Germany).

For preparative co-expression, dense overnight cultures of KA13 cells co-transformed with pMG211_AaLS variants and pACTET_Top36p_GFP36p were used to inoculate 50 mL LB medium (25 µg/mL Cam, 100 µg/mL Amp). Following incubation at 30 °C, 500 ng/mL tetracycline was added when OD_{600} had reached ~0.3 and 200 µM IPTG when OD_{600} ~1.5. After 20 hours incubation at 30 °C, cells were harvested by centrifugation and stored at -20°C before purification by metal affinity chromatography using Ni-NTA agarose and buffers containing 50 mM TRIS, 100 mM NaCl, pH 8.0 and imidazole concentrations between 0-500 mM.

3.7.8.2 Growth of HIV-R10/AaLS co-transformants

Following a protocol described by Wörsdörfer [180], colonies of KA13 cells co-transformed with pMG211_AaLS variants and pACTET_HIV-R10 [179] were used to inoculate 5 mL LB (36 µg/mL Cam, 150 µg/mL Amp, 100 µM salicylate) and incubated overnight at 30 °C. The overnight cultures were diluted to OD_{600} = 0.05 into 5 mL LB (36 µg/mL Cam, 150 µg/mL Amp, 100 µM salicylate, 450 ng/mL tetracycline) in polystyrene tubes (Greiner, Germany) and incubated at 30 °C. OD_{600} values

were determined directly in the tube using an Ultrospec 10 cell density meter (Amersham, GE Life Sciences, USA).

4 Toward sequential enzyme catalysis within a protein container

4.1 Introduction

The ability to target biomacromolecules to the lumen of protein containers stimulated the idea of engineering nanoreactors in which high local concentrations of one or more enzymes are sequestered from bulk solution by a proteinaceous shell. On one hand, such artificial reaction chambers could serve as a means to investigate the metabolic benefit of natural compartments encapsulating sequential enzymes of a metabolic pathway, such as the carboxysome, the propane-1,2-diol utilization and the ethanolamine utilization compartments [103]. On the other hand, engineered nanoreactors might become useful tools in synthetic biology by enabling reaction sequences that generate toxic intermediates inside living cells. Detrimental substances such as aldehydes produced by one enzyme could be contained within the nanoreactor until they are transformed by another enzyme. Another use of nanoreactors might be in accelerating overall metabolic flux or rerouting the flux of a particular metabolite towards a desired pathway by co-encapsulation of the sequentially acting enzymes. In principle, such an outcome could originate from two different and potentially cumulative effects: i) the capsid shell poses a diffusion barrier towards a metabolite and thus increases its residence time within a certain volume around the enzyme active site from which it originated; or ii) high local concentrations of a desired enzyme, combined with very low concentrations of potentially competing enzymes, around the catalytic site that produces a certain metabolite increase the likelihood that this molecule will be processed by the preferred catalyst.

The first example of encapsulating a foreign enzyme in a virus-like particle was presented by Cornelissen, Nolte and co-workers who confined horseradish peroxidase within the empty shell of cowpea chlorotic mottle virus (CCMV) [235]. Following removal of its RNA, the shell was disassembled by increasing pH. Capsid reassembly in presence of enzyme at reduced pH trapped some enzyme molecules within the shell as demonstrated by the peroxidase activity of purified virus-like particles. Using single-molecule fluorescence measurements it was observed that the fluorescent peroxidation product escapes faster from the capsids at increased pH, which is attributable to larger capsid pores. By genetically fusing an α-helical peptide to the N-terminus of the CCMV coat protein, which points towards the inside of the capsid when assembled and a complementary helix to the C-terminus of eGFP, the fluorescent protein could be directed to the capsid's lumen by formation of a charged coiled-coil between the two peptides (Figure 69A) [236]. Using this strategy the same group encapsulated *Candida antarctica* lipase B (CalB) and

found that encapsulation increases its activity relative to the free enzymes – an effect that is also observed when immobilizing this enzyme [122]. The authors attribute the finding that higher loading density lead to a decreased k_{cat} value to reduced substrate availability within the capsid.

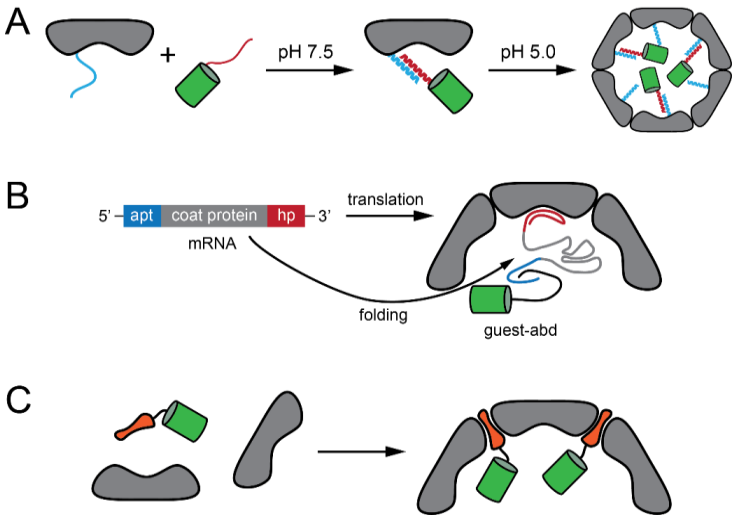


Figure 69. Strategies to target enzymes to the interior of viral capsids. A) Mixing the CCMV coat protein decorated with an α-helix with a guest protein having a complementary helix leads to coiled-coil formation between the two components. Subsequent acidification induces CCMV capsid assembly [236]. B) An mRNA coding for the coat protein of bacteriophage Qβ and containing an aptamer (apt) as well as a hairpin (hp) that interacts with the coat protein, tethers guest proteins fused to an aptamer binding domain (abd) to the interior wall of the virus-like particle [237]. C) Co-expression of a genetic fusion between a protein of interest and the bacteriophage P22 scaffold protein (orange) with the corresponding coat protein affords intact capsids containing one or multiple different enzymes [238].

Francis, Tullman-Ercek and co-workers directed eGFP as well as alkaline phosphatase into the lumen of bacteriophage MS2 by conjugating the guest proteins to DNA or genetically adding a hexadecameric peptide tag composed of glutamates and aspartates. Mixing the proteins decorated with a negatively charged tag with the naturally positively charged MS2 capsid fragments afforded assembled capsids enclosing the guests [239]. It was found that presence of the osmolyte trimethylamine-*N*-oxide enhances capsid formation. Encapsulation reduced the k_{cat} of phosphatase fused to the hexadeca-peptide (1.6 dimers per capsid) by approximately 25 % while leaving the K_m unchanged, suggesting that access of the lumen by the substrate is not limiting. The same strategy was successful in targeting alkaline

phosphatase into MS2 capsids, which bear mutated pores [240]. Introducing one or two mutations changing neutral amino acids to carboxylates increased the apparent K_m of a negatively charged substrate approximately 5- and 2.5-fold, respectively. The authors speculate that the introduction of two additional charges per monomer around the pores situated on the quasi-equivalent 5- fold and 6-fold symmetry axes might have induced conformational changes and larger pores.

Finn and co-workers exploited a high-affinity interaction between an RNA hairpin and the lumenal face of the Qβ bacteriophage coat protein to localize aspartate dipeptidase E and firefly luciferase to its interior [237]. The RNA hairpin was inserted downstream of the gene encoding the Qβ coat protein. Just upstream of the corresponding ribosome binding site the sequence encoding an RNA-aptamer was included. Transcription of this gene carried on a plasmid did not just provide the mRNA used for translation of the coat protein, but at the same time served as tether between capsid coat proteins and guest proteins that had been genetically fused to an aptamer binding domain (Figure 69B). The activity of both enzymes inside and outside the capsid was comparable, with the peptidase experiencing an approximately 3-fold drop in k_{cat}/K_m upon encapsulation. Remarkably, encapsulation protected the peptidase from heat inactivation and proteolysis.

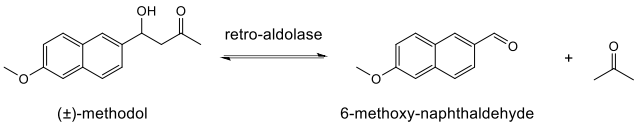
The group of Trevor Douglas developed a strategy to encapsulate guest proteins into the capsid of bacteriophage P22 by genetically fusing it to the capsid's scaffold protein [238]. Co-expression of the guest-scaffold protein fusion with the coat protein affords capsids loaded with the protein of interest (Figure 69C). Capsid extension upon heat treatment to 65°C preserves the complex' integrity. Further heating to 75°C leads to formation of a wiffle-ball structure upon dissociation of the twelve pentamers, leaving behind holes large enough to let the cargo molecules escape. With this strategy a β-glycosidase, an alcohol dehydrogenase and ultimately the enzymes for the cascade reaction discussed extensively in Chapter 1 were located to the lumen of the virus-like particle [34,104,105].

4.2 Targeting enzymes to the lumen of AaLS-13

4.2.1 Retro-aldolase RA95.5-8

Our first attempts to target an enzyme to the lumen of AaLS-13 focused on RA95.5-8 K210M, which catalyzes the reversible retro-aldol cleavage of methodol to 6-methoxy-naphthaldehyde (6MNA) and acetone (Scheme 30) [137]. This enzyme is well-characterized, monomeric, moderately charged and its activity can be assayed directly by the fluorescence or absorption of its product without requiring any coupling enzymes or reagents. Due to the general abundance of surface lysines, many proteins, such as BSA [241], show low basal activity in cleaving methodol. For AaLS-13, however, no such activity above background was detected (Figure 70).

Presumably all six of the lysine residues that are accessible on each monomer are protonated owing to the overall negative charge of the capsid subunit.



Scheme 30. Retro-aldol cleavage of (±)-methodol catalyzed by artificial retro-aldolase RA95.5-8.

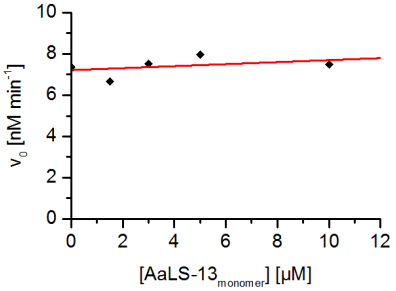


Figure 70. Rate for (±)-methodol cleavage as a function of AaLS-13 concentration. At the $p < 0.05$ significance level, the slope is not significantly different from zero.

Drawing from the successful use of a C-terminal deca-arginine tag (R_{10}) to target GFP and HIV protease to the lumen of AaLS-neg and AaLS-13, respectively [177,179], we introduced this sequence just before the C-terminal His_6-tag of RA95.5-8 K210, the most proficient artificial retro-aldolase available at the outset of this study. The S25C mutation introduced a surface exposed thiol into the otherwise thiol-free enzyme and thus set it up for covalent labeling by a fluorescent dye through a maleimide moiety. Labeling the purified protein, RA95.5-8 S25C/K210M/R_{10} (RA-R10), with an excess of rhodamine-maleimide resulted in protein that was labeled with up to two equivalents of the fluorescent dye as judged by UV-Vis spectroscopy and mass spectroscopy. The modified protein exhibited substantially reduced catalytic activity. Presumably, the severely depressed pK_a of the catalytic lysine 83 renders it an excellent nucleophile towards maleimide. In order to limit covalent modification of this active site residue, fluorophore-labeling was carried out with sub-stoichiometric amounts of reagent in the presence of saturating concentrations of the substrates acetone and 6-methoxy-naphthaldehyde.

RA-R10 labeled with Atto-565-maleimide was mixed with 100 µM purified AaLS-13 capsid in 1:5, 1:7.5, 1:10 and 1:20 ratios relative to the AaLS-13 monomer in encapsulation buffer (50 mM NaP_i, 200 mM NaCl, 5 mM EDTA, pH 8.0) at room temperature. While the majority of the protein precipitated at ratios of 1:5 and 1:7.5, the protein remained soluble at ratios of 1:10 and 1:20. Purifying 1:10 mixtures of RA-R10 and AaLS-13 by size-exclusion chromatography afforded complexes containing 2.7 ± 1.1 molecules of RA-R10 per 180-meric capsid (Figure 71). Re-purification of isolated RA-R10/AaLS-13 complexes by size exclusion chromatography yielded 2.1 ± 0.5 guests per capsid.

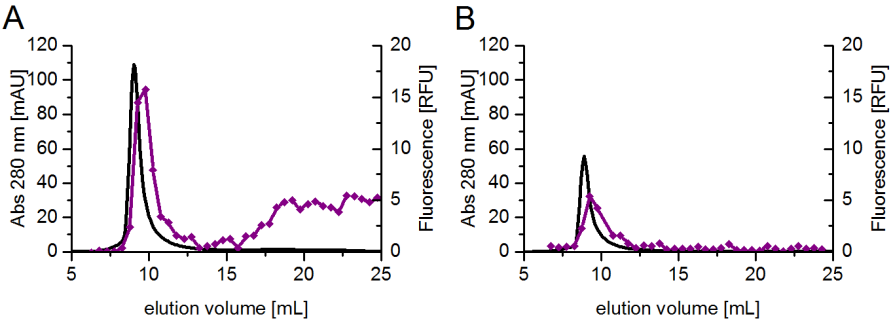
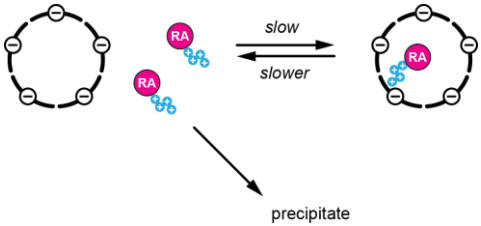


Figure 71. Encapsulation of RA-R10 by AaLS-13 capsids. A) A 1:10 mixture of RA-R10 labeled with ATTO-565. B) Reinjection of purified RA-R10/AaLS-13 complex. The black and purple lines indicate the absorbance at 280 nm and the fluorescence signal (excitation at 565 nm / emission at 592 nm) of each fraction, respectively.

The poor encapsulation efficiency of approximately 15 % contrasts with the ability to retain the encapsulated cargo protein in the capsid during re-purification. In order to obtain reproducible results for the encapsulation and cargo retention, the incubation times between mixing and purification and subsequent re-purification had to be kept constant. Longer incubation times after mixing RA-R10 and AaLS-13 (overnight instead of 1h) increased the fraction of precipitated protein, but at the same time afforded higher loading ratios for the protein remaining in the soluble fraction. The time dependence of encapsulation yields suggests that the determined encapsulation efficiency does not reflect equilibrium. The calculated overall charge of RA-R10 at pH 8.0 is only +4.3, resulting in a low propensity to bind in the lumen of AaLS-13. Competing with encapsulation, RA-R10 can apparently interact with AaLS-13 in a way that ultimately leads to precipitation, a problem that is exacerbated at high concentrations of the enzyme and longer times (Scheme 31). Once the capsid/cargo complex is formed, however, the deca-arginine peptide appears

to interact strongly with the lumenal surface of AaLS-13, presumably by virtue of its localized positive charge, resulting in a high activation barrier for dissociation and thus kinetic stabilization of the capsid-guest complex.



Scheme 31. Encapsulation of RA-R10 by AaLS-13 capsids competes with precipitation.

The mass spectrum of RA-R10 labeled with ATTO-565 shows adducts with m/z +98, a mass corresponding to H_3PO_4. This result suggests that phosphate ions are bound to the deca-arginine tag. It is known that phosphate and guanidinium groups interact strongly, [242] and it is conceivable that phosphate ions in the encapsulation buffer might shield the deca-arginine tag and thus diminish encapsulation. Repetition of the mixing experiment in 50 mM TRIS buffer, however, did not improve encapsulation efficiency (1.3 ± 0.5 RA-R10 per capsid, n = 2).

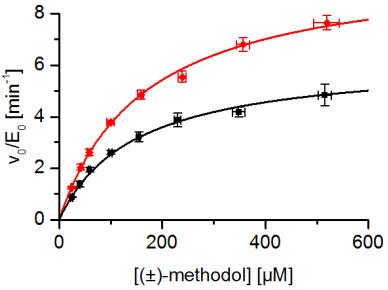


Figure 72. Michaelis-Menten plot for the retro-aldolase activity of free RA-R10-ATTO565 (black squares) and RA-RA10-ATTO565 encapsulated in AaLS-13 (red diamonds, 1.95 ± 0.21 equivalents of RA-R10-ATTO565 per 180-meric AaLS-13 capsid).

The activity of encapsulated RA-R10-ATTO565 is slightly higher compared to free RA-R10-ATTO565 (Figure 72, Table 5). The deca-arginine tag might diminish the activity of the retro-aldolase by interacting with regions of the enzyme involved

in catalysis. When the enzyme is encapsulated in AaLS-13, binding of the deca-arginine tag to the negatively charged capsid wall might alleviate this perturbation, resulting in similar kinetic parameters as for RA95.5-8 K210M.

Table 5. Michaelis-Menten parameters for free and encapsulated RA-R10-ATTO565.

	k_{cat} [min^{-1}]	K_m [μM]	k_{cat}/K_m [M^{-1} min^{-1}]
free labeled RA-R10	5.9 ± 0.6	130 ± 18	(5.9 ± 0.4) · 10^4
encapsulated RA-R10-ATTO-565	10.0 ± 0.6	170 ± 16	(4.6 ± 0.2) · 10^4

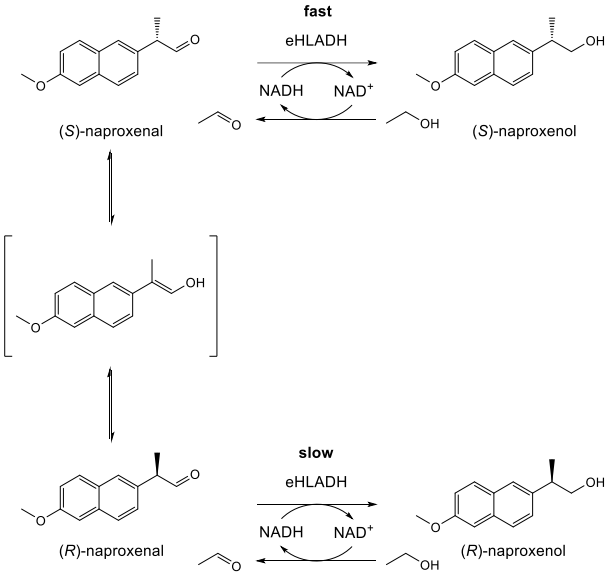
Overall, results from exploiting the deca-arginine tag to encapsulate RA95.5-8 in AaLS-13 are rather disappointing. Encapsulation efficiency falls short of the maximum loading density achieved with supercharged GFP and the unpurified mixtures of RA-R10 and AaLS-13 tend to precipitate over time. In contrast, using supercharged GFP (GFP36+) as an N-terminal fusion tag to direct RA95.5-8 into the AaLS-13 lumen proved to be much more efficient.[1] In addition to retro-aldolase, monomeric and moderately charged enzymes like TEM β-lactamase and a computationally designed Kemp eliminase could be encapsulated quantitatively. In contrast, negatively charged enzymes like cyclohexylamine oxidase and catalase-peroxidase associated with the capsid by at least partially attaching to the exterior surface. Encapsulation of oligomeric enzymes, such as aldehyde dehydrogenase, NADH oxidase and monoamine oxidase, proved difficult since in these cases the majority of the protein precipitated upon mixing the GFP(36+)-enzyme fusion proteins with AaLS-13 capsids.

4.2.2 Alcohol dehydrogenases

The alcohol dehydrogenase from horse liver is a well-characterized homo-dimeric enzyme that exists in two isoforms [243]. The *S* isoenzyme utilizes NADH to reduce the ketone in steroids such as 5β-cholanoic acid 3-one and 5β-androstane-17β-ol-3-one with k_{cat}/K_m values of ~100,000 M^{-1} s^{-1} but is only moderately active towards acetaldehyde (3,000 M^{-1} s^{-1}). Conversely, the *E* isoenzyme (eHLADH) efficiently reduces a wide range of aliphatic and aromatic aldehydes (k_{cat}/K_m ≈ 180,000 M^{-1} s^{-1} for acetaldehyde) [244]. eHLADH has been successfully employed for dynamic kinetic resolution in the synthesis of 2-aryl propanols in the presence of 10 %

[1] This system was designed by Dr. Yusuke Azuma. A report comparing the effectiveness of R10 and GFP(36+) for directing enzymes to the lumen of AaLS-13 has been published [227].

THF or MeCN [245]. For example, racemic naproxenal was reduced at pH 8.0 in 67 % yield and 92 % ee using 10 % MeCN as a co-solvent and 3 % EtOH to recycle NADH (Scheme 32) [246]. Oxidation of the (*S*)-alcohol to the acid gives (*S*)-naproxen, an anti-inflammatory agent. The structural similarity between naproxenal and 6-methoxy-2-naphthaldehyde (6MNA) suggests that the latter might be accepted as a substrate by eHLADH. Furthermore, regeneration of the NADH co-substrate by the small molecule ethanol could be advantageous, since ethanol might diffuse in and out of the capsid much more easily than NADH/NAD^+.



Scheme 32. Dynamic kinetic resolution of naproxenal using eHLADH and an NADH regeneration system [246].

Indeed, eHLADH bearing an N-terminal His-tag [247] catalyzes the reduction of 6MNA with k_{cat} = 4.7 s^{-1} and K_m = 5 µM. This activity was retained in the GFP(36+)-eHLADH fusion protein. Unfortunately, encapsulation of GFP(36+)-eHLADH was not as straightforward as for the monomeric fusion proteins. Mixing GFP(36+)-eHLADH and AaLS-13 at various ratios and ionic strengths primarily afforded precipitate. Small amounts of the GFP(36+)-eHLADH/AaLS-13 guest-host complex could be obtained upon purification of a mixture of 10 µM AaLS-13 and 0.5 equivalents GFP(36+)-eHLADH at *I* = 550 mM (Figure 73). Owing to the minute amounts of purified complex, GFP(36+)-eHLADH concentrations were determined by GFP(36+)

fluorescence. The catalytic parameters of encapsulated GFP(36+)-eHLADH and the free fusion enzyme are virtually identical (Figure 73B and Table 2). At the time when these experiments were performed, the insensitivity of the kinetic parameters to encapsulation was very surprising, since this result implies that diffusion of the bulky co-substrate NADH through the AaLS-13 capsid wall is not rate-limiting. However, rapid diffusion of NADH into and out of the capsid's lumen is in line with the diffusion-limited encapsulation of proteinaceous cargo by AaLS-13 capsids (see Chapter 3).

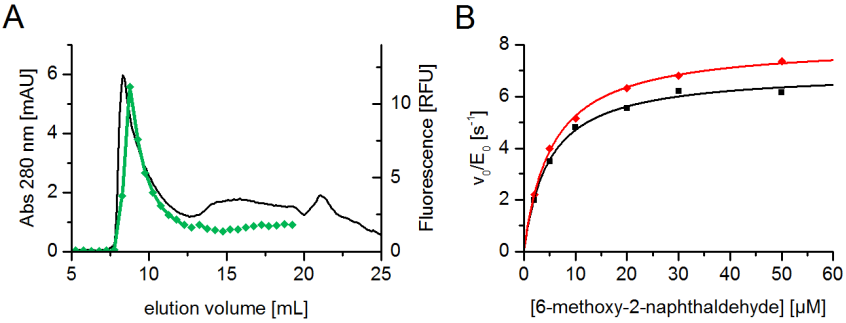


Figure 73. Encapsulation of GFP(36+)-eHLADH by AaLS-13. A) Chromatogram from the separation of a 1:20 mixture of GFP(36+)-eHLADH and AaLS-13 (*I* = 550 mM) on a Superose 6 10/300 GL size exclusion column. The black line indicates on-line absorption at 280 nm, whereas green diamonds reflect the GFP(36+) fluorescence of individual fractions (excitation 488 nm, emission 514 nm). B) Michaelis-Menten plot for reduction of 6-methoxy-2-naphthaldehyde catalyzed by free GFP(36+)-eHLADH (black squares) or purified GFP(36+)-eHLADH/AaLS-13 complex (red diamonds). Reaction conditions are 25 mM HEPES, 100 mM NaCl, pH 7.5, 3 % MeCN, 60 µM NADH, 29 °C.

Table 6. Michaelis-Menten parameters for free GFP(36+)-eHLADH and GFP(36+)-eHLADH encapsulated by AaLS-13. The reported error corresponds to the error of the fit.

	k_{cat} [s^{-1}]	K_m [µM]	k_{cat}/K_m [M^{-1} s^{-1}]
GFP(36+)-eHLADH	7.0 ± 0.2	4.8 ± 0.4	(1.50 ± 0.13) · 10^6
GFP(36+)-eHLADH/AaLS-13 complex	8.1 ± 0.6	5.4 ± 0.2	(1.50 ± 0.06) · 10^6

The low yields of intact GFP(36+)-eHLADH/AaLS-13 complexes, even when mixed at low protein concentrations and increased ionic strength, can be attributed to two effects: i) binding of the homodimeric GFP(36+)-eHLADH to AaLS-13 imposes geometric strain on the eHLADH dimer interface, resulting in its destabilization; or ii) the geometric strain is imposed on the AaLS-13 capsid and destabilizes it. In both cases, the strain might populate protein conformations that have an increased propensity to aggregate. In order to weaken host-guest binding, and thus potentially

reduce any geometric strain, eHLADH constructs were prepared that had either a C-terminal deca-arginine tag (eHLADH-R_{10}) or both an N- and a C-terminal deca-arginine tag (R_{10}-eHLADH-R_{10}). Owing to poor protein expression only tiny amounts of R_{10}-eHLADH-R_{10} could be purified from the supernatant of lysed *E. coli* cells. The nucleotide sequence encoding the N-terminal R_{10}-tag might inhibit protein translation at the mRNA level. Purified eHLADH and eHLADH-R_{10} were mixed with AaLS-13 capsids and the resulting complexes purified by size-exclusion chromatography (Figure 74). Due to lack of a specific chromophore for the enzyme guest, its relative concentration was assayed by measuring the enzyme activity of individual fractions. After mixing AaLS-13 with eHLADH-R_{10}, almost no intact capsid was observed, whereas mixing with eHLADH did not alter capsid structure. In both cases, the highest specific activities were found in fractions containing free enzyme, which eluted around 16 mL. However, the specific activity, normalized by protein concentration (absorption at 280 nm) of the capsid fraction was about 8-fold higher for eHLADH-R_{10}, indicating that the deca-arginine tag could localize eHLADH to the capsids lumen. All in all, however, this system is not suitable for efficiently encapsulating this enzyme in AaLS-13.

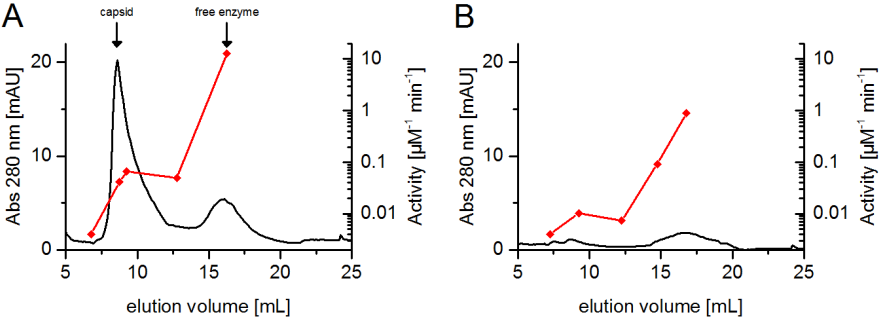


Figure 74. Size-exclusion purification of mixtures of A) eHLADH and B) eHLADH-R_{10} with AaLS-13. Black lines indicate the on-line absorption at 280 nm. Red squares indicate the specific eHLADH activity for reduction of 6MNA in selected fractions. The elution volumes of assembled capsids and free enzyme and/or capsid fragments are indicated by the respective arrow.

Destabilization of the enzyme homodimer by interactions between the GFP36+ fusion partner and AaLS-13 might explain the poor yields of the eHLADH/AaLS-13 complex. To obtain good yield of a functional complex, a thermostable alcohol dehydrogenase was evaluated as an alternative guest. The N-terminus of the tetrameric alcohol dehydrogenase from *Bacillus stearothermophilus*, htADH [248], was genetically fused to GFP36+, expressed in *E. coli*, and purified. Mixing GFP(36+)-htADH

with AaLS-13 capsids in the 1:10 and 1:20 ratios at an ionic strength of 680 mM resulted in complete precipitation, but a 1:30 mixture yielded comparatively little precipitate. Building on other experiments showing that AaLS-neg can encapsulate and retain GFP36+ fusion proteins almost as well as AaLS-13 (see Chapter 3), encapsulation of GFP(36+)-htADH (ratio 1:30) was also attempted with AaLS-neg. For both capsid variants, intact guest-host assemblies could be purified by size-exclusion chromatography as judged by electron microscopy, SDS-PAGE and UV-Vis spectra. Since htADH does not accept 6MNA as a substrate, the oxidation of benzyl alcohol by NAD^+ was monitored. Activities for the free fusion enzyme and the fusion enzyme encapsulated in AaLS-neg or AaLS-13 were similar and about 10-fold lower with respect to k_{cat}/K_m in comparison to the values determined for htADH by Klinman and co-workers (Table 7, Figure 75A) [248]. Due to its low catalytic proficiency at ambient temperatures and the inability to reduce 6MNA, GFP(36+)-htADH is unsuited to test substrate channeling for cascade reactions within the AaLS capsids. As a side note, neither capsid protected the encapsulated fusion protein from proteolysis (Figure 75B).

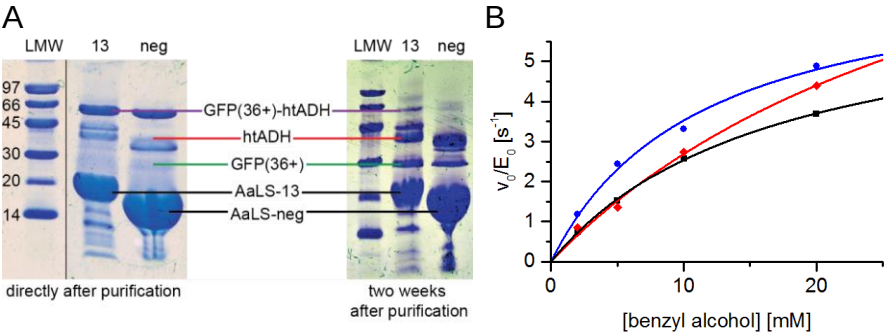


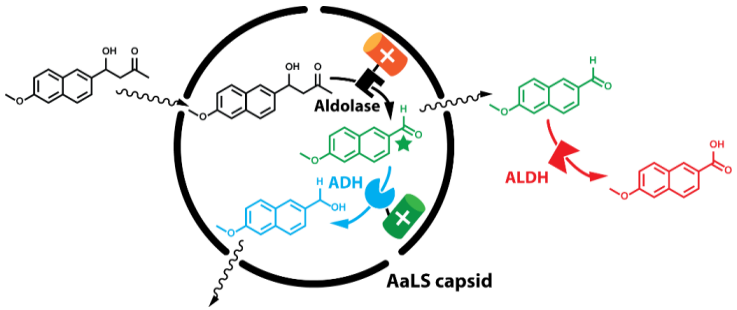
Figure 75. Encapsulation of GFP(36+)-htADH by AaLS-13 and AaLS-neg. A) Comparison of the protein composition directly after purification and two weeks later by SDS-PAGE demonstrates that the fusion protein is proteolyzed despite encapsulation if no protease inhibitor is added (note that lanes unrelated to this experiment were removed from the gel on the left). B) Michaelis-Menten plot for benzyl alcohol oxidation by GFP(36+)-htADH free in solution (black squares) and encapsulated in AaLS-neg (blue circles) or AaLS-13 (red diamonds) measured with 100 nM fusion protein, 10 mM NAD^+ and varying concentrations of benzyl alcohol in 50 mM NaPi, 200 mM NaCl, pH 8.0, 3 % MeCN, 30°C. LMW: low molecular weight marker.

Table 7. Michaelis-Menten parameters for free GFP(36+)-htADH and GFP(36+)-htADH encapsulated by AaLS-neg or AaLS-13. The reported error corresponds to the error of the fit.

	k_{cat} [s^{-1}]	K_m [μM]	k_{cat}/K_m [M^{-1} s^{-1}]
GFP(36+)-htADH	6.8 ± 0.3	16.8 ± 1.2	400 ± 30
GFP(36+)-htADH/AaLS-neg complex	7.5 ± 0.8	11.2 ± 3	670 ± 200
GFP(36+)-htADH/AaLS-13 complex	12 ± 3	35 ± 14	340 ± 160

4.3 Encapsulation of two sequentially acting enzymes in AaLS capsids

RA95.5-8 and alcohol dehydrogenases were chosen for encapsulation in order to be able to perform the cascade reaction depicted in Scheme 33 within the interior of AaLS capsids.



Scheme 33. A model cascade reaction to test substrate channeling in protein containers.

This combination of enzymes would allow testing of substrate channeling using a variety of strategies. All three reactions can be monitored individually using the exact same sample: retro-aldolase (RA) activity is monitored in the absence of nicotinamide nucleosides by addition of methodol, alcohol dehydrogenase (ADH) by addition of 6MNA and NADH, and aldehyde dehydrogenase (ALDH) by addition of 6MNA and NAD^+. Channeling of 6MNA between the active sites of the aldolase and ADH can be assayed by comparing the transient time to reach steady-state for (6-methoxy-2-naphthyl)methanol formation (see Chapter 1) as monitored by NADH consumption in the presence and absence of AaLS-13 capsids and further validated by determining the steady-state concentration of 6MNA by fluorescence. Moreover, the competition method can be employed by addition of external ALDH, which would compete with ADH for the naphthaldehyde intermediate, and subsequent determination of the alcohol/acid ratio in the presence and absence of capsid. In

order to be able to determine the ADH and retro-aldolase concentrations in a purified host-multi-guest complex, the enzymes were respectively fused to GFP36+ and TOP36+, a spectral variant of GFP36+ with a red-shifted absorption and emission spectrum and an orange color.

Because simply mixing GFP(36+)-eHLADH with AaLS-13 at an ionic strength of 350 mM repeatedly yielded precipitate, the insights gained into the encapsulation mechanism (Chapter 3) were utilized to develop an improved encapsulation protocol for this oligomeric enzyme. At low ionic strength, encapsulation of GFP36+ competes with aggregation and ultimately precipitation. It is likely that aggregation is more pronounced for oligomeric enzymes that have multiple covalently appended GFP36+ molecules. Favorable Coulombic interactions presumably increase the lifetime of the initially formed intermediate between such guests and the exterior surface of the capsid, causing concentration-dependent capsid aggregation. Increasing the ionic strength destabilizes this intermediate, reducing the tendency to aggregate, and once the guests have internalized, the ionic strength can be lowered again. In our optimized protocol, guests were therefore mixed with 10 µM AaLS (rather than 100 µM) at an ionic strength of 750 mM (50 mM NaP_i, 600 mM NaCl, 5 mM EDTA, pH 8.0). After incubation for five minutes, the ionic strength was reduced to 350 mM to stabilize the capsid-cargo complex, and the sample was concentrated prior to purification by size-exclusion chromatography.

Following this protocol, both GFP(36+)-eHLADH and TOP(36+)-RA could be co-encapsulated by AaLS-neg but not AaLS-13. However, success depended on the host-guest ratios employed. Whereas encapsulating 30 guests per 180-mer capsid resulted in complete precipitation of protein during the mixing step, reducing the number of guests per capsid to 15 afforded some soluble capsid after size-exclusion. With AaLS-neg, filled capsids (d ≈ 28 nm) were obtained containing full-length fusion proteins. In contrast, AaLS-13 gave largely broken capsids, judging from electron microscopy, which failed to protect the fusion proteins from substantial proteolysis as evidenced by SDS-PAGE analysis (Figure 76).

Failure to obtain stable host-guest complexes with oligomeric GFP(36+)-enzyme fusions and AaLS-13 is likely attributable to the comparatively low stability of AaLS-13 capsids. Although the free energy of intersubunit interactions and thus the free energy of capsid formation is unknown in either case, several observations suggest that the AaLS-13 capsid is less stable than AaLS-neg: i) upon purification from *E.coli* cell lysate, AaLS-neg is isolated exclusively as intact capsids, whereas AaLS-13 yields a mixture of capsid and pentameric subunits; ii) assembled AaLS-neg could be analyzed by native mass spectrometry in a straightforward fashion, whereas this has not yet been possible for AaLS-13; iii) negative-staining TEM never shows broken AaLS-neg capsids, whereas incomplete AaLS-13 capsids are frequently observed; iv)

introduction of two mutations, C52R/C127R, into AaLS-13 renders the capsids kinetically labile (i.e. capsid formation is reversible and assembled capsids are in equilibrium with pentamers; see Chapter 3).

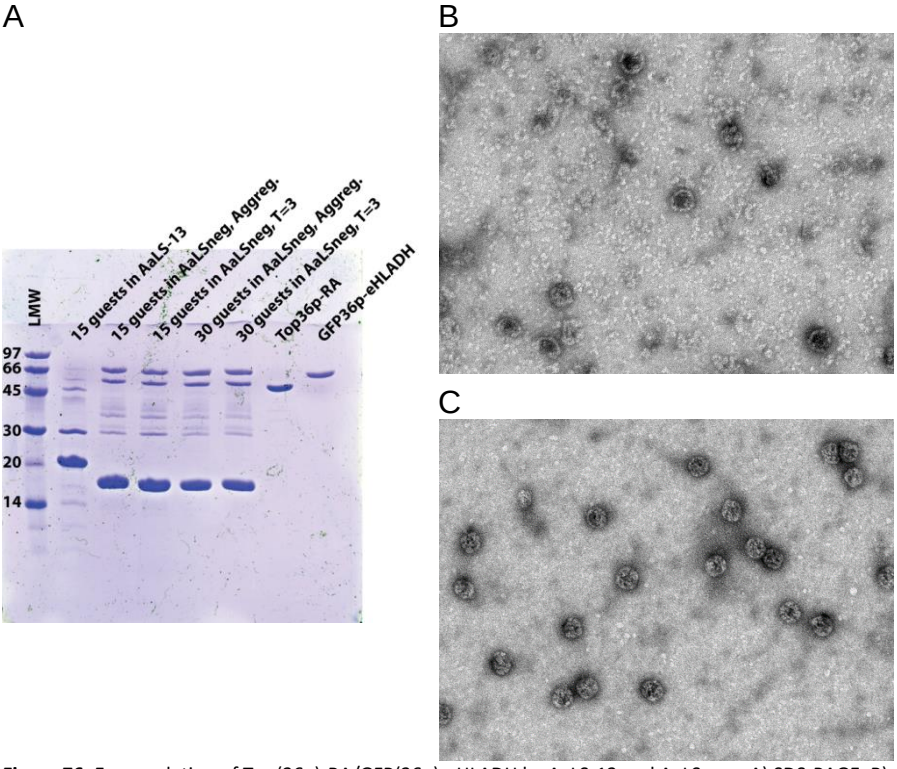


Figure 76. Encapsulation of Top(36+)-RA/GFP(36+)-eHLADH by AaLS-13 and AaLS-neg. A) SDS-PAGE. B) top: EM picture of purified TOP(36+)-RA/GFP(36+)-eHLADH/AaLS-13. C): EM picture of purified TOP(36+)-RA/GFP(36+)-eHLADH/AaLS-neg.

The strong ionic interactions between GFP36+ and the lumenal face of AaLS-13, together with the geometric strain resulting from their spatial orientation induced by the eHLADH dimer, could conceivably disrupt the attractive forces between capsid subunits, driving them apart upon binding GFP(36+)-eHLADH. Contrarily, for AaLS-neg, stronger intersubunit attractions might allow it to withstand this strain. Alternatively, different configurations of the capsid subunit may induce less strain to begin with. Difficulties have been previously encountered with the encapsulation of other oligomeric fusion enzymes by AaLS-13: partial precipitation was observed in the case of a homo-dimeric NADH oxidase and complete precipitation for the

homo-tetrameric enzymes aldehyde dehydrogenase and monoamine oxidase [227]. Nevertheless, such an outcome is not general. For instance, the homo-dimeric carbon fixating protein RuBisCO has been successfully encapsulated by AaLS-13 [249].

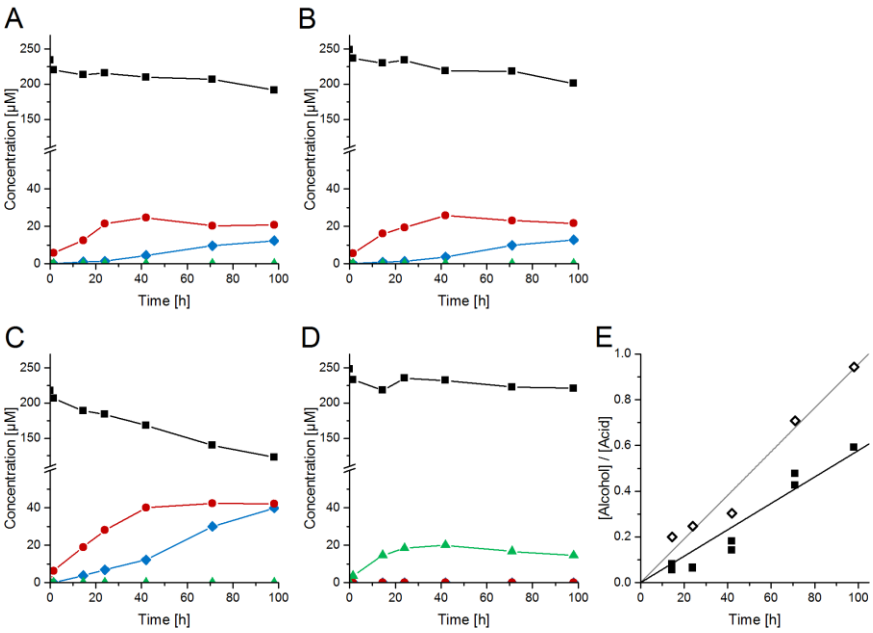


Figure 77. Time-dependent changes in the concentrations of (*R*)-methodol (black squares), 6-methoxy-2-naphthaldehyde (green triangles), (6-methoxy-2-naphthyl)methanol (blue diamonds), and 6-methoxy-2-naphthoic acid (red circles) in the presence of encapsulated and non-encapsulated enzymes. A) Top(36+)-RA/GFP(36+)-eHLADH/AaLS-neg (13 guests per 180-meric capsid) with exogenously added ALDH; B) Top(36+)-RA/GFP(36+)-eHLADH/AaLS-neg (19 guests per capsid) with exogenously added ALDH; C) a mixture of Top(36+)-RA, GFP(36+)-eHLADH and ALDH, and D) TOP(36+)-RA. E) Time dependence of the (6-methoxy-2-naphthyl)methanol : 6-methoxy-2-naphthoic acid ratio for free (open diamonds) and encapsulated enzymes (filled squares). Catalyst concentrations were normalized to 100 nM TOP(36+)-RA in all samples with a TOP(36+)-RA/GFP(36+)-eHLADH ratio of 1:2.

Although the attraction between GFP36+ and capsid is weaker for AaLS-neg than for AaLS-13, as demonstrated by slower guest exchange for the latter (see Chapter 3), the interactions apparently suffice to yield thermodynamically stable guest-host complexes. Moreover, the smaller size of AaLS-neg capsids compared to AaLS-13 (d = 28 nm and 38 nm, respectively) might benefit enzymatic cascade reactions through increased molecular crowding. Since AaLS-neg and AaLS-13 retain

positively charged cargo proteins by close interaction with the lumenal surface, the ability to encapsulate cargo is limited by the lumenal surface area. Since the ratio of volume to surface of a sphere scales as $V/A = 1/3 \cdot r$, the enzyme density in a filled 180-meric AaLS-neg capsid (19.6 mM for 36 guests, r_{in}=9 nm) is about twice as high as in a filled 360-meric AaLS-13 capsid (10.4 mM for 72 guests, r_{in} = 14 nm).

The partitioning of 6-methoxy-2-naphthaldehyde to (6-methoxy-2-naphthyl)methanol and 6-methoxy-2-naphthoic acid via reduction by alcohol dehydrogenase or oxidation by aldehyde dehydrogenase, respectively, is readily monitored by UPLC. The time course of their formation after addition of methodol to a ternary complex of Top(36+)-RA/GFP(36+)-eHLADH/AaLS-neg in the presence of free ALDH (Scheme 33) is shown in Figure 77.

Consistent with the findings obtained by following the retro-aldol cleavage of methodol by UV-VIS spectroscopy [227], the activity of TOP(36+)-RA is considerably impeded by encapsulation and product inhibition [137] if the naphthaldehyde is not removed from solution (Figure 77D). Turnover of TOP(36+)-RA necessitates long reaction times, during which ALDH, but not GFP(36+)-eHLADH, inactivates. This inactivation is apparent from the leveling off of the naphthoic acid concentration after 40 h. Despite having 1 mM DTT in the reaction mixture, the active-site cysteine of ALDH might be irreversibly oxidized over 2 days at room temperature. Slow inactivation of ALDH is further reflected in the time-dependent ratio of alcohol and acid product.

Channeling of the naphthaldehyde intermediate to the alcohol dehydrogenase in this system would be evident from skewing the product distribution in favor of the alcohol product in presence of capsid. Overall, however, our data show that co-encapsulation of TOP(36+)-RA and GFP(36+)-eHLADH does not appear to favor channeling. The slightly lower conversion seen for encapsulated GFP(36+)-eHLADH might be caused by the somewhat reduced activity of this enzyme when encapsulated. In light of the very low reaction rates, reduced availability of the negatively charged cofactor NADH within the negatively charged capsid is unlikely to explain the absence of channeling.

The influence of encapsulation on TOP(36+)-RA and GFP(36+)-eHLADH activity was evaluated by addition of methodol or naphthaldehyde/NADH, respectively, to the Top(36+)-RA/GFP(36+)-eHLADH/AaLS-neg complex and to a mixture of the free enzymes. The retro-aldol reaction clearly limits the overall cascade, since naphthaldehyde is rapidly reduced in both cases (Figure 78).

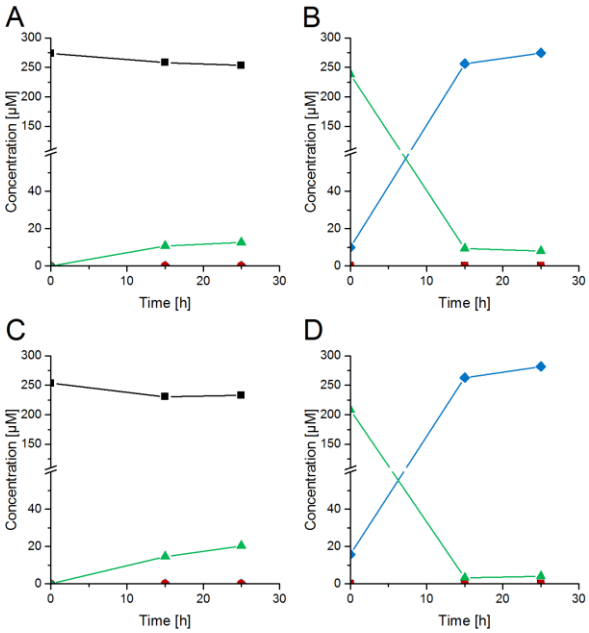


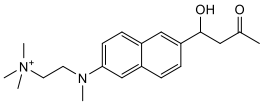
Figure 78. Individual enzyme activities for the Top(36+)-RA/GFP(36+)-eHLADH/AaLS-neg complex (A, B) or a mixture of free Top(36+)-RA and GFP(36+)-eHLADH (C, D). Retro-aldolase activity was monitored following addition of 200 µM (*R*)-methodol (A, C). Alcohol dehydrogenase activity was monitored following addition of 200 µM 6-methoxy-2-naphthaldehyde and 1 mM NADH (B, D). Concentrations of (*R*)-methodol (black squares), (6-methoxy-2-naphthyl)methanol (blue diamonds), 6-methoxy-2-naphthoic acid (red circles) and 6-methoxy-2-naphthaldehyde (green triangles) were determined by UPLC using a calibration curve.

4.4 Conclusion and outlook

A variety of enzymes can be encapsulated by AaLS-neg or AaLS-13 by covalent fusion to positively supercharged green fluorescent protein. Despite stronger electrostatic interactions between targeting tag and the lumen of AaLS-13, the higher stability of the AaLS-neg capsid might render this variant more suitable for encapsulation of oligomeric guests. The slightly lower kinetic stability of GFP36+/AaLS-neg capsids resulting from the somewhat lower charge at the shell interior is not critical on the timescales relevant to investigations of enzyme cascade reactions. On the contrary, the reduced charge of AaLS-neg might limit intermediate and permanent aggregation alike and thus result in higher yields for the ternary complexes. Ideally, however, preparation of encapsulated enzyme mixtures with both capsids

would be desirable, as this would allow evaluation of the influence of capsid size and thus catalyst concentration on substrate channeling.

The rapidity with which AaLS-neg and AaLS-13 encapsulate and release proteinaceous guests, together with the wild-type rates of some internalized enzymes, suggest that the capsid walls do not limit diffusion of small molecules. Substrate channeling could thus emerge either through dense co-clustering of enzymes within the capsid or by retaining the intermediate via favorable interactions with the charged capsid shell. To that end, a positively charged derivative of methodol might be employed (Scheme 34) [250].



Scheme 34. Positively charged analog of methodol used as a substrate for an ultrahigh-throughput droplet-based microfluidic retro-aldolase screening assay [250].

The results obtained in preliminary experiments on the aldolase/alcohol dehydrogenase complex suggest that this system will be useful for investigating cascade reactions. Enzyme activities can be monitored individually by varying substrates while maintaining identical enzyme complexes in all cases. Competition experiments can be conducted by addition of an exogenous enzyme and determination of the product distribution by UPLC. Whether the bulk steady state aldehyde concentration is reduced by encapsulation can be determined by fluorescence. Such an effect would be potentially useful for synthetic applications in vivo and in vitro alike. Lowered aldehyde concentrations would be expected to limit cytotoxicity by reducing nonspecific Schiff-base formation with lysine residues as well as inhibition of synthetic aldol reactions (see Chapter 2). The moderate substrate specificity of artificial aldolases results in formation of an unproductive Schiff base adduct between the more electrophilic acceptor aldehyde and the catalytic lysine, blocking the active site. Removal of the aldehyde by oxidation or reduction would ameliorate this problem.

However, the activities of the enzymes employed in the prototypic system presented in this chapter are too slow, not only to effect substrate channeling, but also to conveniently monitor the reactions. Despite encapsulation, the flexible linker of the fusion proteins is still proteolyzed – an effect exacerbated under the reaction conditions. Addition of protease inhibitors has to be carefully balanced so that they do not interfere with product analysis by UPLC. More effective enzymes would not only reduce the time needed to perform the reaction, the resulting increased product concentrations would also render the protease inhibitors less of a problem.

Fusing the supercharged fluorescent protein tags to the N-terminus of RA95.5-8 was chosen for convenience. However, as it turns out, the N-terminus is not surface exposed–at least after removal of the methionine by endogenous *E. coli* aminopeptidases. C-Terminal modifications are much less detrimental to enzyme activity in this system than N-terminal substitutions. In the meantime, Obexer et al. have increased the activity of RA95.5-8 by 30-fold by directed evolution (k_{cat} = 11 s^{-1} and k_{cat}/K_m = 34,000 M^{-1} s^{-1}), accompanied by an improvement in stability and decreased inhibition by 6-methoxy-2-naphthaldehyde [250]. These properties make the enzyme a potentially much better candidate for encapsulation. The characterization of active site mutants demonstrated that variants with 5 to 100-fold reduced k_{cat} could be easily accessed by single-point mutations if necessary. The availability of enzymes with a broad range of catalytic parameters in an otherwise identical system might be useful for dissecting the factors influencing substrate channeling [34].

Since high proficiency of the downstream enzyme is an even more important prerequisite for metabolite channeling, catalysts that reduce or oxidize 2-naphthylaldehydes more rapidly would be beneficial. To that end, the recently identified benzyl alcohol and aldehyde dehydrogenases from the *Pseudomonas putida* strain CSV86 are potentially attractive candidates. CSV86 is known to metabolize naphthalene and methylnaphthalenes. Recently, an alcohol dehydrogenase and aldehyde dehydrogenase were purified from this bacterium that respectively reduce and oxidize 2-naphthylaldehyde with k_{cat}/K_m values of ~$7.4 \cdot 10^7$ M^{-1} s^{-1} and ~$1.5 \cdot 10^7$ M^{-1} s^{-1} via hydride transfer to or from nicotinamide nucleotides [251]. By comparing the published N-terminal peptide sequences that had been determined by Edman degradation, with the sequence of putative dehydrogenases found in the genome [252], the codon-optimized DNA sequences encoding for the two candidate catalysts were synthesized and will soon be evaluated for their suitability as catalysts for cascade reactions.

4.5 Materials and methods

4.5.1 Molecular biology

E. coli XL1blue was used as a host for all cloning steps. Plasmids containing confirmed gene sequences were transformed into BL21 gold DE3 for protein production. Genes were amplified by PCR using Phusion High-Fidelity DNA polymerase and the primers mentioned in the text. The sequences of the employed oligonucleotides are listed in Table 8. PCR products were purified by agarose gel electrophoresis. Following digestion, the plasmid backbone was dephosphorylated with Antarctic phosphatase prior to purification by agarose gel electrophoresis. Phusion High-Fidelity DNA polymerase, restriction enzymes, Antarctic phosphatase and T4 ligase were purchased from New England BioLabs (Ipswich, USA). Oligonucleotides were synthesized by Microsynth AG (Balgach, Switzerland). Sequences of all modified genes were confirmed by DNA sequencing performed at Microsynth AG (Balgach, Switzerland).

RA-R10. The gene encoding RA95.5-8 K210M was amplified from plasmid pET29b_RA95.5-8_K210M using primers RA_fw and RA-R10_rv. The PCR product and the pET29b(+) plasmid were digested with *Nde*I and *Xho*I and subsequently ligated to give plasmid pET29b_RA-R10. The S25C mutation was introduced into pET29b-RA-R10 using a ligation-during-amplification procedure with forward primer FW_S25C and reverse primer RV_S25C, yielding pET29b_RA(S25C)-R10.

R10-eHLADH-R10. Plasmid pRSETb-EqADH-E was a generous gift from Francesca Paradisi [247]. In order to remove an internal *Nde*I restriction site, gene fragments amplified using primers T7 and eHLADH_xNdeI_rv, or eHLADH_xNdeI_fw and eHLADH_SpeI_rv, were assembled by PCR using primers T7 and eHLADH_SpeI_rv. The resulting product was digested with *Nde*I and *Spe*I and ligated into accordingly digested pET29b_RA(S25C)-R10 to give pET29b_LEADER-eHLADH-R10. Oligonucleotides NdeI-R10-SacI_fw and NdeI-R10-SacI_rv were annealed and then ligated into plasmid pET29b_LEADER-eHLADH-R10 that had been digested with *Nde*I and *Sac*I to yield pET29b_R10-eHLADH-R10.

***His6-GFP(36+)-eHLADH*.** The gene encoding the E isoform of horse liver alcohol dehydrogenase was amplified from pRSETb-EqADH-E using primers XhoI_eHLADH_fw and eHLADH-GSS_SpeI_rv. The resulting DNA product was digested with *Xho*I and *Spe*I, and subsequently ligated into accordingly digested pACYC_His6-GFP36p-RA, a generous gift from Dr. Yusuke Azuma [227], to afford pACYC_His6-GFP36p-eHLADH.

***GFP(36+)-eHLADH-His6*.** In order to facilitate purification, the His6-tag was moved to the C-terminus of the fusion protein by amplifying pACYC_His6-GFP36p-eHLADH with the NdeI_GFP36p_fw/eHLADH_xNdeI_rv and eHLADH_xNdeI_fw/eHLADH_SacI_His6_SpeI_rv primer pairs. The two products were assembled by PCR

using primers NdeI_GFP36p_fw and eHLADH_SacI_His6_SpeI_rv. The assembled gene was digested with *Nde*I and *Spe*I and ligated into accordingly digested pACYC_His6-GFP36p-eHLADH to yield pACYC_GFP36p-eHLADH-His6.

***GFP(36+)-htADH-His6*.** The plasmid encoding a high-temperature alcohol dehydrogenase from *Bacillus stearothermophilus* (htADH) was a generous gift from Prof. Judith Klinman [248]. The htADH gene was amplified using primers XhoI_htADH_fw and htADH_SacI_rv and the resulting product digested with *Xho*I and *Sac*I. The restricted DNA was ligated into accordingly digested pACYC_GFP36p-eHLADH-His6 to yield pACYC_GFP36p-htADH-His6.

***TOP(36+)-RA-His6*.** Primers NdeI_GFP36p_fw and linker_BamHI_rv were used to amplify the gene encoding TOP36+ from pACYC_His6-TOP36p (Chapter 3). Primers linker_BamHI_fw and RA_SacI_rv were used to amplify RA from pACYC_His6-GFP36p-RA. Both genes were assembled using primers NdeI_GFP36p_fw and RA_SacI_rv, and the resulting gene product was digested with *Nde*I and *Sac*I before ligation into accordingly digested pACYC_His6-GFP36p-eHLADH to afford pACYC_Top36p-RA-His6. An additional *Bam*HI restriction site between Top36+ and the flexible linker allows facile exchange of the N-terminal tag.

***BADH-TOP(36+)-His6*.** ATG Biosynthetics (Germany) designed and synthesized the *badh-linker-his6* gene and inserted it into pET17b (Novagen/Merck, Germany) to yield pET17b_BADH-GSS-His6. This plasmid was designed such that the gene encoding TOP(36+) could be amplified from plasmid pACYC_TOP36p with primers BamHI_GFP36p_fw and GFP36p_SacI_rv, digested with *Bam*HI and *Sac*I, and ligated into this vector to yield pET17b_BADH-TOP36p-His6. BADH is the putative 2-naphthylmethanol dehydrogenase from *P. putida* CSV 86 [251,252].

***BZDH-His6*.** ATG Biosynthetics (Germany) designed and synthesized the *bzdh-his6* gene and inserted it into pET17b to yield pET17b_BZDH-His6. BZDH is the putative 2-naphthylaldehyde dehydrogenase from *P. putida* CSV 86 [251,252].

Table 8. Oligonucleotides used for molecular biology.

Name	Sequence
BamHI_GFP36p_fw	TCCGGCGGATCCATGGCTAGCAAAGGTGAACG
eHLADH_SacI_His6_SpeI_rv	TCAGCTGACTAGTCAGTGGTGGTGATGATGGTGAGCTCCAAACGTCAGGATGGTACG
eHLADH_SpeI_rv	TCAGCTGACTAGTCAAAACGTCAGGATGGTACGGATACTCTCTCC
eHLADH_xNdeI_fw	TGTCATGCTGTCAAGAAGCATACGGTGTGAGCGTCATTGTGG
eHLADH_xNdeI_rv	TGCTTCTTGACAGCATGACAAGGCAGTCACC
eHLADH-GSS_SpeI_rv	CGTCTACGACTAGTGGAGCTGCCAAACGTCAGGATGGTACGGATACTCTCTCC
GFP36p-SacI_rv	GATGGTGAGCTCCCTTGTAGCGTTCGTCGCGTCC
htADH_SacI_rv	GATGGTGAGCTCCTTTATCTTCTAACGTTAAAACTACG
linker_BamHI_fw	CGAACGTGGTGGATCCGGGGGCAGCGGCGGTTCGGGCG
linker_BamHI_rv	CGCTGCCCCCGGATCCACCACGTTCGAGCGCCATGGAGC
NdeI_GFP36p_fw	GATATACATATGGCTAGCAAAGGTGAACG
NdeI_R10_SacI_fw	TATGCGTAGACGACGCCGTCGGCGACGTCGACGTGGGAGCT
NdeI_R10_SacI_rv	CCCACGTCGACGTCGCCGACGGCGTCGTCTACGCA
RA_fw	TGATAACATATGCCGCGTTATCTGAAAGG
RA_rv	TTATCACTCGAGTGATCCCTCGATTAATTCTTT
RA-R10_rv	GGTGCTCGAGACGTCGACGTCGCCGACGGCGTCGTCTACGACTAGTTGATCCCTCGATTAAT
RA-SacI_rv	ATGGTGAGCTCCTGATCCCTCGATTAATTCTTTAATC
S25C_fw	CGTCCACGCGTGCCGCCAACGTCCGATTATTTC
S25C_rv	GACGTTGGCGGCACGCGTGGACGCTAGGACGGCGC
T7	TAATACGACTCACTATAGG
T7term	CCGCTGAGCAATAACTAGCA
XhoI_eHLADH_fw	GGCGGCTCGAGCATGAGCACAGCAGGAAAAGTAATAAAATGC
XhoI_htADH_fw	GGCGGCTCGAGCATGAAAGCCGCTGTAGTTGAAC

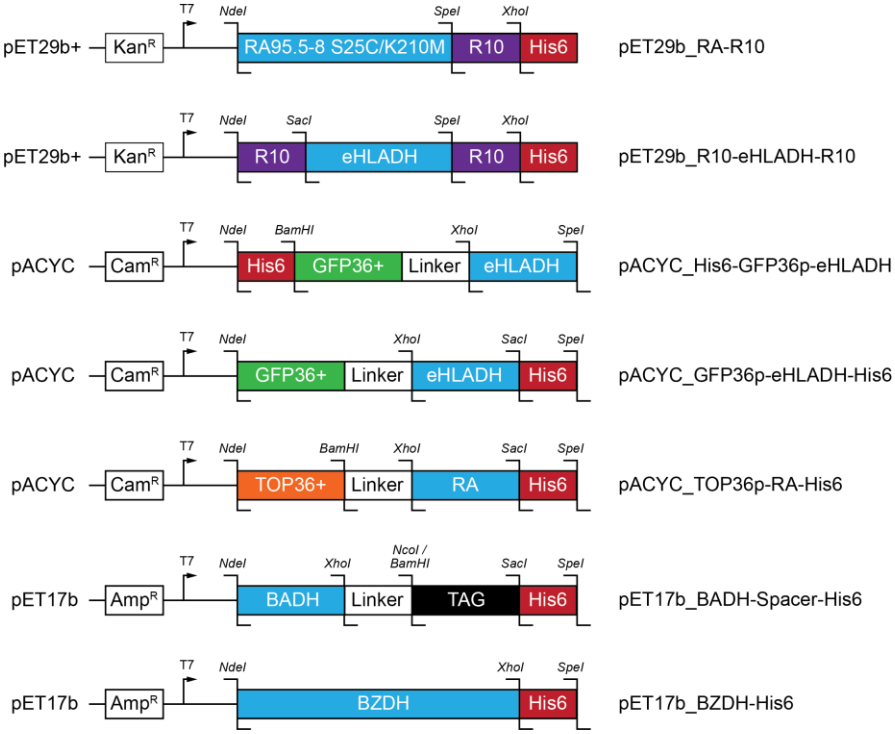


Figure 79. Structure of genes and plasmids created in this work, including from left to right: plasmid type, resistance genes, promoter and restriction sites.

4.5.2 Protein sequences

RA-R10 (RA95.5-8 S25C/K210M R10)

MPRYLKGWLEDVVQLSLRRPSVHACRQRPIISLNERILEFNKRNITAIIAYYTRK
SPSGLDVERDPIEYAKYMERYAVGLSIKTEEKYFNGSYEMLRKIASSVSIPILMN
DFIVKESQIDDAYNLGADTVLLIVNILTERELESLLEYARSYGMEPLILINDEND
LDIALRIGARFIVIFSMNFETGEINKENQRKLISMIPSNVVKVAMLDISERNEIE
ELRKLGVNAFLISSSLMRNPEKIKELIEGSTSRRRRRRRRRRLEHHHHHH*

Mutations within the RA gene are highlighted in red.

LEADER-eHLADH-R10

MRGSHHHHHHGMASMTGGQQMGRDLYDDDDKDPSSMSTAGKVIKCKAAVLWEEKK
PFSIEEVEVAPPKAHEVRIKMVATGICRSDDHVVSGTLVTPLPVIAGHEAAGIVE
SIGEGVTTVRPGDKVIPLFTPQCGKCRVCKHPEGNFCLKNDLSMPRGTMQDGTSR
FTCRGKPIHHFLGTSTFSQYTVVDEISVAKIDAASPLEKVCLIGCGFSTGYGSAV

KVAKVTQGSTCAVFGLGGVGLSVIMGCKAAGAARIIGVDINKDKFAKAKEVGATE
CVNPQDYKKPIQEVLTEMSNGGVDFSFEVIGRLDTMVTALSCCQEAYGVSVIVGV
PPDSQNLSMNPMLLLSGRTWKGAIFGGFKSKDSVPKLVADFMAKKFALDPLITHV
LPFEKINEGFDLLRSGESIRTILTFGSSTSRRRRRRRRRLEHHHHHH*

R10-eHLADH-R10

MRRRRRRRRRGSSMSTAGKVIKCKAAVLWEEKKPFSIEEVEVAPPKAHEVRIKM
VATGICRSDDHVVSGTLVTPLPVIAGHEAAGIVESIGEGVTTVRPGDKVIPLFTP
QCGKCRVCKHPEGNFCLKNDLSMPRGTMQDGTSRFTCRGKPIHHFLGTSTFSQYT
VVDEISVAKIDAASPLEKVCLIGCGFSTGYGSAVKVAKVTQGSTCAVFGLGGVGL
SVIMGCKAAGAARIIGVDINKDKFAKAKEVGATECVNPQDYKKPIQEVLTEMSNG
GVDFSFEVIGRLDTMVTALSCCQEAYGVSVIVGVPPDSQNLSMNPMLLLSGRTWK
GAIFGGFKSKDSVPKLVADFMAKKFALDPLITHVLPFEKINEGFDLLRSGESIRT
ILTFGSSTSRRRRRRRRRLEHHHHHH*

His6-GFP(36+)-eHLADH

MHHHHHHGSGMASKGERLFRGKVPILVELKGDVNGHKFSVRGKGKGDATRGKLTL
KFICTTGKLPVPWPTLVTTLTYGVQCFSRYPKHMKRHDFFKSAMPKGYVQERTIS
FKKDGKYKTRAEVKFEGRTLVNRIKLKGRDFKEKGNILGHKLRYNFNSHKVYITA
DKRKNGIKAKFKIRHNVKDGSVQLADHYQQNTPIGRGPVLLPRNHYLSTRSKLSK
DPKEKRDHMVLLEFVTAAGIKHGRDERYKSGGSMALERGGSGGSGGSGGSGGSSM
STAGKVIKCKAAVLWEEKKPFSIEEVEVAPPKAHEVRIKMVATGICRSDDHVVSG
TLVTPLPVIAGHEAAGIVESIGEGVTTVRPGDKVIPLFTPQCGKCRVCKHPEGNF
CLKNDLSMPRGTMQDGTSRFTCRGKPIHHFLGTSTFSQYTVVDEISVAKIDAASP
LEKVCLIGCGFSTGYGSAVKVAKVTQGSTCAVFGLGGVGLSVIMGCKAAGAARII
GVDINKDKFAKAKEVGATECVNPQDYKKPIQEVLTEMSNGGVDFSFEVIGRLDTM
VTALSCCQEAYGVSVIVGVPPDSQNLSMNPMLLLSGRTWKGAIFGGFKSKDSVPK
LVADFMAKKFALDPLITHVLPFEKINEGFDLLRSGESIRTILTF*

The sequence SGGSMALER highlighted in magenta has been unintentionally included in the GFP(36+)-RA fusion protein [227] because it is part of the open reading frame of pACYC-GFPpos [181] from which GFP(36+) had been subcloned.

GFP(36+)-eHLADH-His6

MASKGERLFRGKVPILVELKGDVNGHKFSVRGKGKGDATRGKLTLKFICTTGKLP
VPWPTLVTTLTYGVQCFSRYPKHMKRHDFFKSAMPKGYVQERTISFKKDGKYKTR
AEVKFEGRTLVNRIKLKGRDFKEKGNILGHKLRYNFNSHKVYITADKRKNGIKAK
FKIRHNVKDGSVQLADHYQQNTPIGRGPVLLPRNHYLSTRSKLSKDPKEKRDHMV
LLEFVTAAGIKHGRDERYKSGGSMALERGGSGGSGGSGGSGGSSMSTAGKVIKCK
AAVLWEEKKPFSIEEVEVAPPKAHEVRIKMVATGICRSDDHVVSGTLVTPLPVIA
GHEAAGIVESIGEGVTTVRPGDKVIPLFTPQCGKCRVCKHPEGNFCLKNDLSMPR
GTMQDGTSRFTCRGKPIHHFLGTSTFSQYTVVDEISVAKIDAASPLEKVCLIGCG
FSTGYGSAVKVAKVTQGSTCAVFGLGGVGLSVIMGCKAAGAARIIGVDINKDKFA
KAKEVGATECVNPQDYKKPIQEVLTEMSNGGVDFSFEVIGRLDTMVTALSCCQEA

YGVSVIVGVPPDSQNLSMNPMLLLSGRTWKGAIFGGFKSKDSVPKLVADFMAKKF
ALDPLITHVLPFEKINEGFDLLRSGESIRTILTFGAHHHHHH*

GFP(36+)-htADH-His6

MASKGERLFRGKVPILVELKGDVNGHKFSVRGKGKGDATRGKLTLKFICTTGKLP
VPWPTLVTTLTYGVQCFSRYPKHMKRHDFFKSAMPKGYVQERTISFKKDGKYKTR
AEVKFEGRTLVNRIKLKGRDFKEKGNILGHKLRYNFNSHKVYITADKRKNGIKAK
FKIRHNVKDGSVQLADHYQQNTPIGRGPVLLPRNHYLSTRSKLSKDPKEKRDHMV
LLEFVTAAGIKHGRDERYKSGGSMALERGGSGGSGGSGGSGGSSMKAAVVEQFKE
PLKIKEVEKPTISYGEVLVRIKACGVCHTDLHAAHGDWPVKPKLPLIPGHEGVGI
VEEVGPGVTHLKVGDRVGIPWLYSACGHCDYCLSGQETLCEHQKNAGYSVDGGYA
EYCRAAADYVVKIPDNLSFEEAAPIFCAGVTTYKALKVTGAKPGEWVAIYGIGGL
GHVAVQYAKAMGLNVVAVDIGDEKLELAKELGADLVVNPLKEDAAKFMKEKVGGV
HAAVVTAVSKPAFQSAYNSIRRGGACVLVGLPPEEMPIPIFDTVLNGIKIIGSIV
GTRKDLQEALQFAAEGKVKTIIEVQPLEKINEVFDRMLKGQINGRVVLTLEDKGA
HHHHHH*

TOP(36+)-RA-His6

MASKGERLFRGKVPILVELKGDVNGHKFSVRGKGKGDATRGKLTLKFICTTGKLP
VPWPTLVTTLGYGVQCFARYPKHMRRHDFFKSAMPKGYVQERTISFKKDGKYKTR
AEVKFEGRTLVNRIKLKGRDFKEKGNILGHKLRYNFNSHKVYITADKRKNGIKAK
FKIRHNVKDGSVQLADHYQQNTPIGRGPVLLPRNHYLSYRSKLSKDPKEKRDHMV
LLEFVTAAGIKHGRDERYKSGGSMALERGGSGGSGGSGGSGGSSMPRYLKGWLED
VVQLSLRRPSVHASRQRPIISLNERILEFNKRNITAIIAYYTRKSPSGLDVERDP
IEYAKYMERYAVGLSIKTEEKYFNGSYEMLRKIASSVSIPILMNDFIVKESQIDD
AYNLGADTVLLIVNILTERELESLLEYARSYGMEPLILINDENDLDIALRIGARF
IVIFSMNFETGEINKENQRKLISMIPSNVVKVAMLDISERNEIEELRKLGVNAFL
ISSSLMRNPEKIKELIEGSGAHHHHHH
Mutations distinguishing TOP36+ from GFP36+ are highlighted in red.

BADH-GSS-His6

MTTPLFRPIQAAVTRGKGAPFVLEQAGIRAPQGDEVLVRVVATGMCHTDMIVRDQ
YYPVPLPAVLGHEGSGIVEAVGPLVQNLAVGDHVVMTYGYCGQCLACDAGQAAYC
QDFFGRNFSGAGPQGQHALQDAQGEMLNDHFFAQSSFATFALGRENNTVRVPKEA
PIELLGPLGCGIQTGAGAVINSLKVTPGSAFAAFGGGAVGLSAVLAAQVAGAATI
IAIDVVPSRLALALELGATHVVNSREVDPVTTIREITGGGVQFALESTGRPEVLR
QAVDALGSRGALGVVGAPPLGTTAQFDVNDLLLGGKVIRGIVEGDSVPKKFIPEL
VNLYLQGRFAFDKLVRFYDFEQINQAAEDSEKGLTLKPIIRIQKGSSGGSGGSGG
SGGSMASKGGAHHHHHH*

BZDH-His6

METLLPPQLWNERLFTGSWVPGELPVTTVTSPGSGQPLGTLANAGTKQVEQSAQE
AAKAQRNWYALPYDERASVLRKAAAIAEQYQAEIADWIVRESGSTQSKASFESSI

TIKSLHEASALPSRSQGEVLPSTAGRLSIARRRPLGVIGVISPFNFPLYLAMRAV
APALALGNAVVLKPDPRTAVCGGHVIARIFELAGLPNGVLHVLPGGAEAGAALTA
SPSVAMIQFTGSTPAGRKVGEAAGRHLKKVSLELGGKNSLIVLDDADLDLAIANA
TWGVYLHQGQICMSTGRLLVQRGIYAAFVERLVAKARSLTVGDPASGQVHLGPLI
NAAQRDNALRIVEAARAAGATLATGGSADGLFFQPTVLSDVQRDNPAFHEEIFGP
VAVVVPFDSDEQAVELANNTDYGLSAAILSRDTGRALRLGEQLRTGLLHINDQTV
NDEVINPFGGVGASGNGTSIGGAANWEEFTQWQWLTLKSEAPAYPLGSSHHHHHH
*

4.5.3 Protein production, purification and labeling

4.5.3.1 RA-R10, eHLADH-R10, R10-eHLADH-R10

Erlenmeyer flasks containing 400 mL sterile LB medium and 30 µg/mL kanamycin were inoculated with dense overnight cultures of *E. coli* XL1blue transformed with pET29b_RA(S25C)-R10, pET29b_LEADER-eHLADH-R10, or pET29b_R10-eHLADH-R10 and incubated at 37 °C and 230 rpm. Following induction with 100 µM IPTG, the temperature was reduced to 30 °C, and 16-20 h later cells were harvested by centrifugation (4000 g, 4 °C, 20 min). The cell pellet was stored at -20 °C.

Cell pellets were thawed and suspended in lysis buffer (25 mM HEPES, 1.6 M NaCl, 50 mM $MgCl_2$, pH 7.5) supplemented with a commercial protease inhibitor cocktail (cOmplete protease inhibitor cocktail, Roche, Switzerland), 1 mg/mL lysozyme, and a spatula tip each of DNase I and RNase A. Following incubation at room temperature for 1 hour, cells were sonicated and the lysate cleared by centrifugation (14,000 g, 4 °C, 30 min). From here on, all solutions were kept at 4 °C. The supernatant was loaded on 4 mL of Ni-NTA agarose resin equilibrated with lysis buffer. Following washing with lysis buffer, and lysis buffer supplemented with 20 and 40 mM imidazole, the protein was eluted with elution buffer (25 mM HEPES, 0.6 M NaCl, 50 mM $MgCl_2$, 500 mM imidazole, pH 7.5). Protein purity was checked by SDS-PAGE with Coomasie blue staining. Protein-containing fractions were combined and buffer exchanged to 25 mM HEPES, 0.6 M NaCl, 5 mM EDTA, pH 7.5 by repeated concentration and dilution using centrifugal ultrafiltration units (30 kDa MWCO, Amicon Ultra, Merck Millipore). A protease inhibitor cocktail (Sigma-Aldrich, USA) was added to the concentrated protein sample prior to storage at 4 °C.

In order to label RA(S25C)-R10, the purified protein was treated with 10 mM tris(2-carboxyethyl)phosphine) (TCEP) for 10 min to reduce any disulfide bridges. After removing excess TCEP on a PD-10 desalting column (GE Healthcare), between 50-100 µM RA(S25C)-R10 was reacted with 1.0 equivalent of Atto-565-maleimide in 2.4 mL labeling buffer containing 1 M acetone and 200 µM 6-methoxy-2-naphthaldehyde in order to inhibit reaction between the maleimide and the reactive lysine at position 83. After incubation in the dark at room temperature for 2 h, unreacted reagent was quenched by addition of 10 mM β-mercaptoethanol. Removal

of low molecular weight compounds including unreacted dyes using a PD-10 desalting column yielded RA-R10-ATTO565. The degree of labeling was determined by absorbance using extinction coefficients $\varepsilon_{dye,565nm}$ = 120,000 M^{-1} cm^{-1}, $\varepsilon_{dye,280nm}$ = 19,200 M^{-1} cm^{-1} and $\varepsilon_{RA,280nm}$ = 21,890 M^{-1} cm^{-1}. The degree of labeling was typically between 20-40 %. Covalent modification of RA-R10 was assessed by electron spray ionization mass spectrometry (ESI-MS): unlabeled protein, 31,499.4 Da (calc. 31,499.3 Da for construct lacking the N-terminal methionine); labeled protein: 31,499.8 Da and 32,133.8 Da (calc. 32,132.6 Da for construct lacking the N-terminal methionine). The labeled RA-R10 protein (RA-R10-ATTO565) was supplemented with a protease inhibitor cocktail (Sigma-Aldrich, USA) and stored at 4 °C.

4.5.3.2 GFP(36+)-/TOP(36+) fusion proteins

Erlenmeyer flasks containing 400 mL sterile LB medium and 36 µg/mL chloramphenicol were inoculated with dense overnight cultures of *E. coli* XL1blue transformed with pACYC_Fusion-Protein and incubated at 37 °C and 230 rpm. Following induction with 100 µM IPTG, the temperature was reduced to 25 °C and 20-24 h later cells were harvested by centrifugation (4000 g, 4 °C, 20 min). The cell pellet was stored at -20 °C.

Cell pellets were thawed and suspended in lysis buffer (50 mM NaP_i, 2 M NaCl, pH 7.4) supplemented with a protease inhibitor cocktail (cOmplete protease inhibitor cocktail tablets, Roche, Switzerland), 1 mg/mL lysozyme and a spatula tip each of DNase I and RNase A. Following incubation at room temperature for 1 hour, cells were sonicated and the lysate cleared by centrifugation (14,000 g, 4 °C, 30 min). From here on, all solutions were kept at 4 °C. The supernatant was loaded on 4 mL of Ni-NTA agarose resin equilibrated with lysis buffer. Following washing with lysis buffer, and lysis buffer supplemented with 20 mM imidazole, the protein was eluted with lysis buffer supplemented with 250 mM imidazole. Protein purity was checked by SDS-PAGE with Coomasie blue staining. Protein-containing fractions were combined and concentrated using centrifugal ultrafiltration units (30 kDa MWCO, Amicon Ultra, Merck Millipore). A MonoS HR 10/10 A column (GE Healthcare) was equilibrated with 50 mM TRIS, 500 mM NaCl, pH 7.4 at 4°C. Prior to loading on the cation exchange column, the concentrated protein was diluted with 3 volumes of 50 mM TRIS, pH 7.4 to reduce the ionic strength. Proteins were eluted with a linear gradient from 50 mM TRIS, 500 mM NaCl, pH 7.4 to 50 mM TRIS, 1500 mM NaCl, pH 7.4 (TRIS buffer was used because NaP_i crystallizes at high sodium concentrations and low temperatures). Fractions containing the protein of interest were combined, concentrated, buffer exchanged to 50 mM NaP_i, 600 mM NaCl, pH 8.0, supplemented with a protease inhibitor cocktail and 20 % glycerol, and stored at -20 °C.

The advantage of the GFP(36+)-guest-His6 construct manifests itself in facile removal of proteolyzed fusion protein, i.e. either GFP(36+) or the guest-His6. During the metal affinity chromatography, GFP(36+) is simply removed by washing, whereas guest-His6 is easily removed by washing the cation exchange column with low salt buffer.

4.5.4 Preparation, purification and characterization of guest-host complexes between fusion proteins and AaLS capsids

In a typical procedure, assembled AaLS-13 capsids are added to a solution of guest fusion proteins at various concentrations in 50 mM NaP_i, 200 mM NaCl, 5 mM EDTA, pH 8.0 so that the final concentration of capsids is 100 µM with respect to AaLS-13 monomers in a total volume of 500 µL. In order to prevent irreversible aggregation of the oligomeric guests, the procedure was modified by increasing the ionic strength (50 mM NaP_i, 600 mM NaCl, 5 mM EDTA, pH 8.0), reducing the AaLS-13 concentration to 10 µM, and increasing the volume to 5 mL. In the latter case, the mixture was concentrated to 500 µL using a centrifugal ultrafiltration unit (30 kDa MWCO, Amicon Ultra, Merck Millipore), diluted to 5 mL with standard buffer (50 mM NaP_i, 600 mM NaCl, 5 mM EDTA, pH 8.0), and concentrated to 500 µL. The concentrated samples were filtered and purified by size exclusion chromatography using a Superose 6 10/300 GL (GE Healthcare) column equilibrated with standard buffer at room temperature. The chromatogram was recorded by measuring online absorbance at 260, 280, 488 and 513 nm. The elution of fluorescent proteins or proteins modified with a fluorescent probe was monitored by measuring the fluorescence of individual fractions using a plate reader (Varioskan, Thermo Electron Corporation). In the case of fluorophore-labeled RA-R10, emission at 592 nm upon excitation at 565 nm was recorded; in the case of fusions to GFP36+, the sample was excited at 488 nm and emission at 509 nm recorded. Fractions containing capsids were combined, concentrated, and stored at room temperature. Protein composition was analyzed by SDS-PAGE, negative-staining electron microscopy, and UV-VIS spectroscopy, as described in Chapter 3.

4.5.5 Determination of enzyme kinetic parameters

4.5.5.1 Retro-aldolase

The rate of cleavage of (±)-methodol into 6-methoxy-naphthaldehyde and acetone was measured as previously described [137,138]. Briefly, varying concentrations of (±)-methodol (added from a stock solution in MeCN) in 25 mM HEPES, 100 mM NaCl, pH 7.5, 2.7 % MeCN were incubated in a sealed quartz cuvette at 29 °C. (±)-Methodol concentrations were determined by absorption at 330 nm (ε_{330nm} = 1,420 M^{-1} cm^{-1}). Reaction was initiated by addition of protein sample containing the

retro-aldolase at a final concentration of 50 or 100 nM. 6-Methoxy-2-naphthaldehyde formation was monitored spectrophotometrically by an increase in absorption at 350 nm ($\Delta\varepsilon_{350nm}$ = 5,970 M^{-1} cm^{-1}) on a Perkin Elmer Lambda 35 UV-vis spectrometer. Steady-state kinetic parameters were obtained by fitting the rate of product formation normalized by enzyme concentration as a function of (±)-methodol concentration to the Michaelis-Menten equation using Origin 9.1 (OriginLab Corporation).

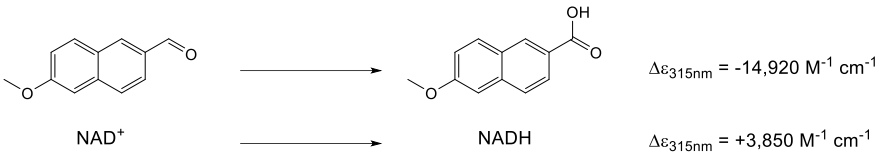
4.5.5.2 Alcohol dehydrogenase

***Horse liver alcohol dehydrogenase*.** The rate of reduction of 6-methoxy-2-naphthaldehyde to (6-methoxy-2-naphthyl)methanol was followed spectrophotometrically by a decrease in absorption at 340 nm. The total change in absorption ($\Delta\varepsilon_{340nm}$ = 14,300 M^{-1} cm^{-1}) originates from the reduced absorption of NAD^+ relative to NADH ($\Delta\varepsilon_{340nm}$ = 6,220 M^{-1} cm^{-1}) as well as (6-methoxy-2-naphthyl)methanol compared to 6-methoxy-2-naphthaldehyde ($\Delta\varepsilon_{340nm}$ = 8,080 M^{-1} cm^{-1}). Varying concentrations of 6-methoxy-naphthaldehyde (added from a stock solution in MeCN) were incubated in 25 mM HEPES, 100 mM NaCl, pH 7.5, 3.0 % MeCN, 60 µM NADH at 29 °C. Reaction was started by addition of 1 or 10 nM encapsulated or pure alcohol dehydrogenase fusion protein.

High temperature alcohol dehydrogenase. The rate of oxidation of benzyl alcohol to benzaldehyde by NAD^+ was followed spectrophotometrically by an increase in absorption at 340 nm ($\Delta\varepsilon_{340nm}$ = 6,220 M^{-1} cm^{-1}). Varying concentrations of benzyl alcohol (added from a stock solution in MeCN) were incubated in 50 mM NaP_i, 200 mM NaCl, pH 8.0, 3.0 % MeCN, 10 mM NAD^+ at 30 °C. Reaction was started by addition of 100 nM encapsulated or pure alcohol dehydrogenase fusion protein.

4.5.5.3 Aldehyde dehydrogenase

Purified *E. coli* aldehyde dehydrogenase, ALDH [253], was a generous gift from Dr. Yusuke Azuma [227]. The rate of oxidation of acetaldehyde or 6-methoxy-2-naphthaldehyde to acetic or naphthoic acid, respectively, was monitored spectrophotometrically at 340 or 315 nm, respectively. The decrease in absorption for naphthoic acid compared to naphthaldehyde at 340 nm is almost identical to the increase in absorption for NADH compared to NAD^+. For this reason, following the reaction at 340 nm is impractical. Instead, the difference extinction coefficients for naphthoic acid and naphthaldehyde as well as NADH and NAD+ were determined at 315 nm (Scheme 35). The total change in absorption is $\Delta\varepsilon_{315nm}$ = -11,070 M^{-1} cm^{-1}.



Scheme 35. Change in absorption at 315 nm for the redox pairs 6-methoxy-2-napthoic acid / 6-methoxy-2-naphthaldehyde and NAD^+/NADH.

Varying concentrations of acetaldehyde (added from an aqueous stock solution) were incubated in 50 mM NaP_i, pH 8.0, 1 mM NAD^+, 1 mM DTT at 25 °C. Reaction was started by addition of 50 nM aldehyde dehydrogenase. In the case of 6-methoxy-2-naphthaldehyde, varying concentrations of substrate (added from a stock solution in MeCN) were incubated in 50 mM NaP_i, pH 8.0, 1 mM NAD^+, 1 mM DTT, 2.7 % MeCN at 25 °C. Reaction was started by addition of 0.1 or 1 µM aldehyde dehydrogenase. In the case of 6-methoxy-2-naphthaldehyde, quantitative conversion to 6-methoxy-2-naphthoic acid was confirmed by UPLC using a 6-methoxy-2-naphthoic acid standard (ABCR, Karlsruhe).

Whereas dehydrogenation of acetaldehyde obeyed Michaelis-Menten kinetics, oxidation of 6-methoxy-2-naphthaldehyde was subject to substrate inhibition, as frequently encountered for aldehyde dehydrogenases [251,254,255] (Table 9 and Figure 80).

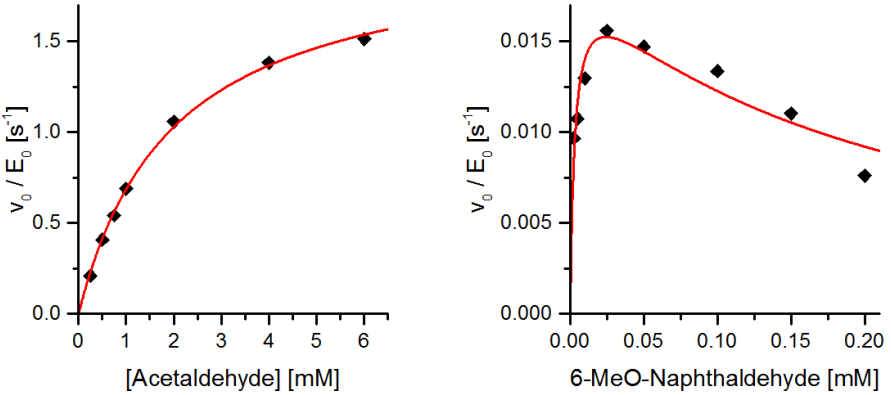


Figure 80. Initial rate plotted as a function of substrate concentration for *E. coli* aldehyde dehydrogenase in presence of 1mM NAD^+.

Table 9. Kinetic parameters for oxidation of aldehydes by *E. coli* aldehyde dehydrogenase. The error is the error of the fit.

Substrate	k_{cat} [s^{-1}]	K_m [mM]	k_{cat}/K_m [M^{-1} s^{-1}]	K_I [µM]
acetaldehyde	2.04 ± 0.04	2.0 ± 0.1	1,000 ± 100	n.d.
6-MeO-2-naphthal-dehyde	0.019 ± 0.002	0.003 ± 0.001	6,000 ± 2,000	190 ± 60

4.5.5.4 Cascade reactions

Substrate depletion and product formation were monitored by UPLC. A 50 µL sample was mixed with 50 µL MeCN and precipitate removed by centrifugation. The supernatant was injected onto a C18 column (XBridge, C18, 2.5 µm, 4.6 x 75 mm, Waters) and the sample eluted with a linear gradient from 5 % to 95 % MeCN (0.08 % TFA) in water (0.1 % TFA), recording the absorption of the eluate at 240 nm. Using this method, methodol and the derived 6-methoxy-2-naphthyl alcohol, aldehyde and acid were baseline resolved (Figure 81).

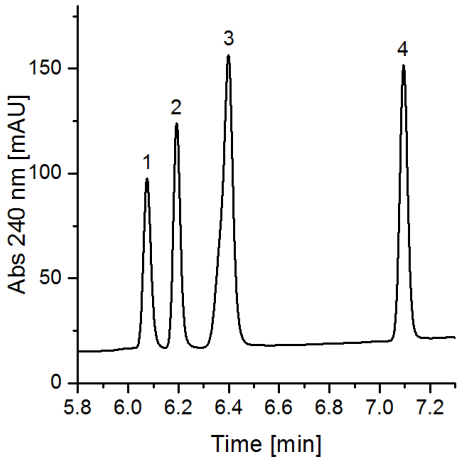


Figure 81. Chromatogram of a sample containing 50 µM each of (6-methoxy-2-naphthyl)methanol (1), (*R*)-methodol (2), 6-methoxy-2-naphthoic acid (3) and 6-methoxy-2-naphthaldehyde (4).

Ternary complexes of TOP(36+)-RA, GFP(36+)-eHLADH and AaLS-neg were added to a solution containing 200 µM (*R*)-methodol, 1 mM NAD^+, 1mM NADH, 1 mM DTT and 50 µM ALDH in 50 mM NaP_i, 5 mM EDTA, pH 8.0, 2.7 % MeCN at room

temperature. The amount of complex added was adjusted to yield a final concentration of GFP(36+)-eHLADH of 0.2 µM. To compensate for the approximately 240-fold difference in k_{cat}/K_m between GFP(36+)-eHLADH and ALDH, the solution contained a corresponding multiple of ALDH with respect to the GFP(36+)-eHLADH concentration. The time course of the reaction was determined by analyzing the metabolite concentration of aliquots using the method described above. A sample containing all enzymes at identical concentrations but without AaLS-neg served as a control.

5 Targeting cargo to the lumen of wild-type lumazine synthase

5.1 Introduction

Two decades after it had been demonstrated that *Bacillus subtilis* lumazine synthase/riboflavin synthase is an icosahedral T=1 capsid constructed from 60 β-subunits that encapsulate a α-subunit trimer [206], Ban and co-workers discovered another family of bacterial protein containers and christened them encapsulins [173]. The crystal structure of the encapsulin from the hyperthermophilic bacterium *Thermotoga maritima* shows an icosahedral T=1 capsid assembled from sixty monomers. Closer inspection revealed short peptides bound to quasi-equivalent clefts in the lumen of the shell (Figure 82). The peptides were identified as the C-terminus (GGDLGIRK) of a ferritin-like protein. Subsequently, encapsulins from other organisms were found to enclose different guests, including a dye-decolorizing peroxidase. While the purpose of encapsulating these proteins in unknown, the conserved hydrophobic pockets on the lumenal surface of the encapsulins suggest a common mechanism for targeting them to the inside of the protein shell. Analysis of a homologous ferritin-like protein, which forms a pentamer of dimers, indicated that the arrangement of the C-termini matches the positions of the symmetrically related binding pockets in the container. Electron microscopic analysis of the *Brevibacterium linens* encapsulin showed electron density consistent with a trimer of dimers with D3 symmetry within the protein shell [173]. Building on these findings, Cornelissen and co-workers demonstrated that the C-terminus of the *B. linens* dye-decolorizing peroxidase can be used as a tag to direct a fluorescent protein into *B. linens* encapsulins [216].

Interestingly, in bacterial microcompartments such as the propane-1,2-diol utilization compartment, guest proteins are located to the lumenal side of the shell protein by virtue of N-terminal rather than C-terminal peptides [256,257]. Fusions of these peptides to the N-terminus have been successfully employed to target foreign proteins to bacterial microcompartments [258,259].

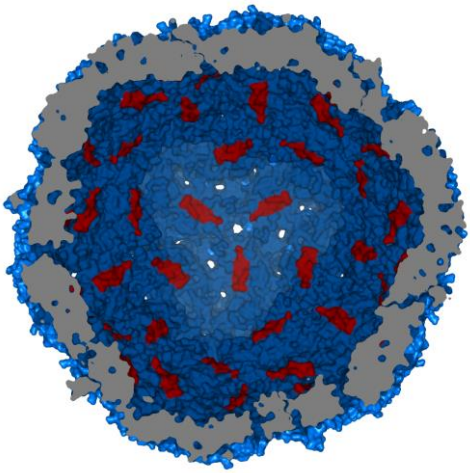


Figure 82. Cutaway view of a *T. maritima* encapsulin, oriented along a three-fold symmetry axis, shows the symmetric arrangement of bound targeting peptides (red) [173].

Although no bound peptides were found in the crystal structures of heterologously expressed dodecahedral lumazine synthase capsids [260], the guest riboflavin synthase might be directed to the lumenal space by a similar mechanism. In this context, the pseudo-C_3-symmetric C-terminal domain found in all riboflavin synthases might mediate multivalent interactions with binding sites around the C_3 axis on the interior of lumazine synthase capsids (Figure 83). The three intertwined helices found at the C-terminus of riboflavin synthases in crystal structures of the enzymes from *Escherichia coli* (PDB 1IED), *Schizosaccharomyces pombe* (PDB 1KZL) and *Brucella abortus* (PDB 4E0F) are reminiscent of homotrimeric coiled-coils [260–262].

The characteristic motif of homotrimeric coiled-coils, HHxxHxx, in which H signifies an amino acid with a hydrophobic sidechain, most commonly leucine or isoleucine, appears in both the *E. coli* and the *A. aeolicus* sequence (Table 10). However, in contrast to typical trimeric coiled-coils, the close packing of hydrophobic residues is preserved over two helical turns only. In addition, the C-terminal helices in riboflavin synthase are arranged at angles of 46-58° relative to one another, rather than the 16° angle usually adopted [263], and consequently diverge towards the C-terminus (Figure 84). This arrangement is reflected in the increasing distances between Cα atoms (e.g., 6.9 ± 0.3 Å for V191, 7.3 ± 0.3 Å for V195, and 9.7 ± 0.8 Å for L199 versus the 6.9 ± 0.1 Å distance preserved across the full length of the tri-

meric coiled-coil from chicken cartilage matrix protein [263]). The divergent α-helices of *E. coli* riboflavin synthase are followed by a short peptide, the structure of which has not been resolved. This C-terminal peptide is likely to be the recognition motif that directs this protein to the lumen of lumazine synthase. The relative orientation of the C-terminal helices might position the three peptides in such a way that each can interact with putative binding sites around the C_3 symmetry axes in the capsid.

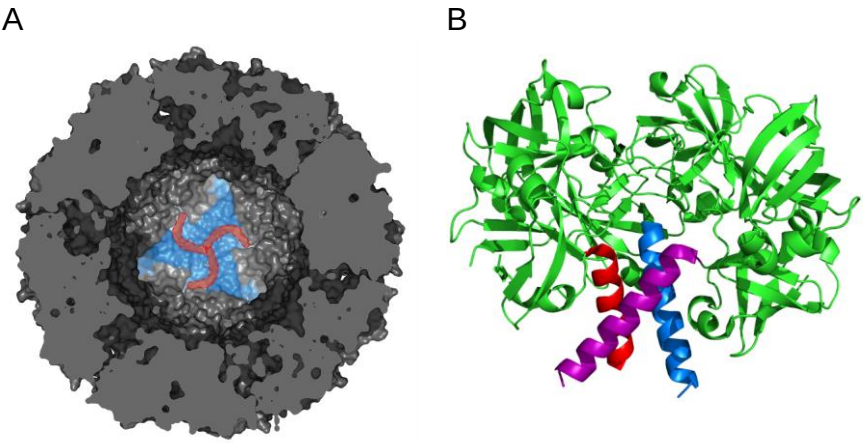


Figure 83. Symmetry elements that could facilitate interactions between the lumenal surface of AaLS and the C-termini of homo-trimeric riboflavin synthase. A) Cutaway view along a three-fold symmetry axis of an AaLS capsid. The clefts that constitute the enzyme active site are highlighted in red. B) Structure of the *E. coli* riboflavin synthase homo-trimer. The C-terminal α-helices of each monomer are shown in different colors [260]. Figures were generated with PyMOL from structures 1HQK and 1I8D deposited in the protein database.

Table 10. Sequence alignment of the C-termini of *E. coli* and *A. aeolicus* riboflavin synthases and a trimeric coiled-coil from chicken cartilage matrix protein (CMP). In the trimeric coiled-coil repetition motif H signifies a hydrophobic amino acid and x any amino acid.

PDB	Protein	Organism	Res	Protein sequence
1I8D	Riboflavin synthase	*E. coli*	185	DPQTQ AVVDT VERVL AAREN AMNQP GTEA
n/a	Riboflavin synthase	*A. aeolicus*	181	DILGK YVINY LNKLK KKEDI FKEFL KW
1AQ5	CMP	*G. gallus*	460	CKSIV KFQTK VEELI NTLQQ KLEAV AKRIE ALEN
		helical wheel assignment		..... .gabc defga bcdef gabcd efg
		trimeric coiled-coil repetition motif		.xxHx xHHxx HxxHH xxHxx HHxxH xx

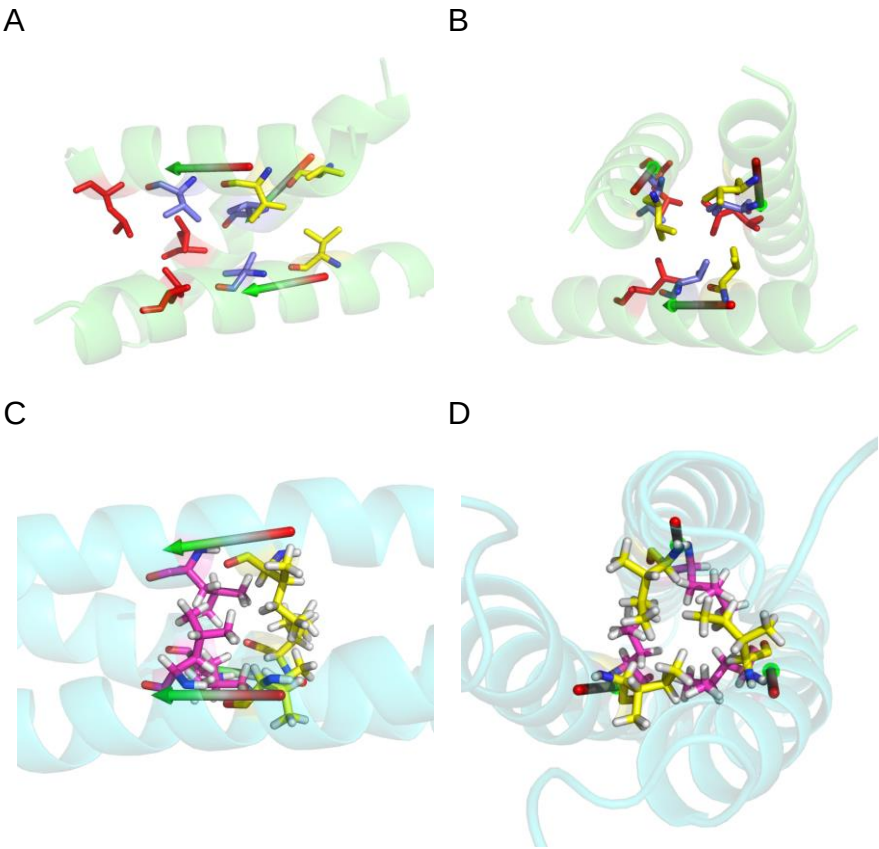


Figure 84. Homo-trimeric pseudo coiled-coil of *E. coli* riboflavin synthase (top, PDB 1I8D) and the coiled-coil of *Gallus gallus* cartilage matrix protein (bottom, PDB 1AQ5). A/B) C-terminal helices of *E. coli* riboflavin synthase with V191 in yellow, V195 in purple, and L199 in red. C/D) The coiled-coil found in chicken cartilage matrix protein with I474 in yellow and L477 in magenta. The arrows indicate the directionality of the α-helices. Figures were generated with PyMOL.

In order to determine which structural features of *A. aeolicus* riboflavin synthase are required for encapsulation by lumazine synthase, several fusion protein constructs of superfolder green fluorescent protein (sfGFP) were prepared and characterized [264].

5.2 Results

Riboflavin synthase (AaRS) and His6-tagged variants of lumazine synthase (AaLS) from *Aquifex aeolicus* were heterologously co-expressed in *E. coli*. The samples were purified by metal affinity chromatography and characterized by analytical gel electrophoresis under reducing conditions. As shown in Figure 85, AaRS co-purified with AaLS-wt and AaLS-neg, but not with AaLS-13. These results are consistent with the findings of Bacher et al. who co-purified identical complexes from *Bacillus subtilis* [265].

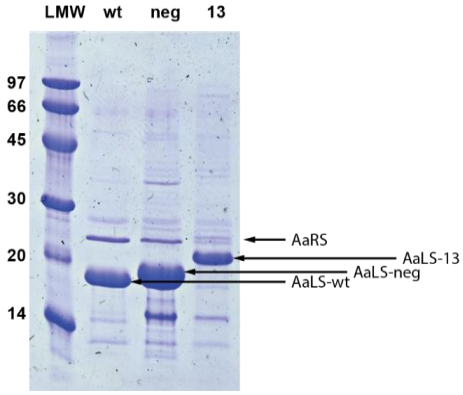


Figure 85. SDS-PAGE analysis of purified *E. coli* cell lysate following co-expression of AaRS and AaLS variants.

Encouraged by these results, genes encoding three different fusion proteins were prepared in which the C-terminus of AaRS was fused to sfGFP. In sfGFP-L1-helix and sfGFP-L2-helix the last 29 amino acids of AaRS were appended to the C-terminus of sfGFP, connected by a GSG or a GSGGSG linker, respectively. For sfGFP-coil, the terminal dodeca-peptide, corresponding to the amino acids not resolved in the crystal structure of *E. coli* riboflavin synthase, was fused via a GSG linker. These three constructs plus sfGFP as a negative control were co-expressed with AaLS-wt and AaLS-neg in *E. coli* and purified by metal affinity and size-exclusion chromatography (Figure 86). AaLS-wt elutes as a mixture of peaks although analysis by electron microscopy only indicates closed-shell particles with a diameter of ~20 nm. If AaLS-wt is co-expressed with sfGFP bearing various C-terminal peptides from AaRS, the fractions between 7-15 mL, corresponding to capsids, are fluorescent, whereas they are not if AaLS-wt is co-expressed with untagged sfGFP. Similar results were obtained with both fusion proteins. These results indicate that tagged sfGFP variants are associated with the capsid, but sfGFP alone does bind to AaLS-

wt capsids. In contrast, none of the fusion proteins coproduced with AaLS-neg capsids afforded fluorescent particles (fraction eluting around 10 mL).

These results suggest that the presumably unstructured C-terminal dodecapeptide of AaRS suffices to localize proteins to the lumen of AaLS-wt. The helical segment seems not to be required for encapsulation. Identical elution volumes for sfGFP-L1-helix and sfGFP-coil indicate that the helix does not induce trimerization of sfGFP. From the UV spectrum of sfGFP-tag/AaLS-wt complexes an average of 0.44 ± 0.04 sfGFP per 60-meric AaLS-wt capsid was determined (Figure 87A). Given that GFP36+ (derived from sfGFP) has an extinction coefficient, $\varepsilon_{488nm,}$ of 36,600 M^{-1} cm^{-1} [227], it is possible that the reported value for sfGFP of 83,300 M^{-1} cm^{-1} is overestimated by a factor of approximately two. In this case, roughly one sfGFP molecule would be contained in an AaLS capsid, which is what would be expected given the space available within the lumen (Figure 87B). Electron microscopy images indicate that sfGFP encapsulation is not accompanied by morphological changes in the capsid (Figure 87C and D).

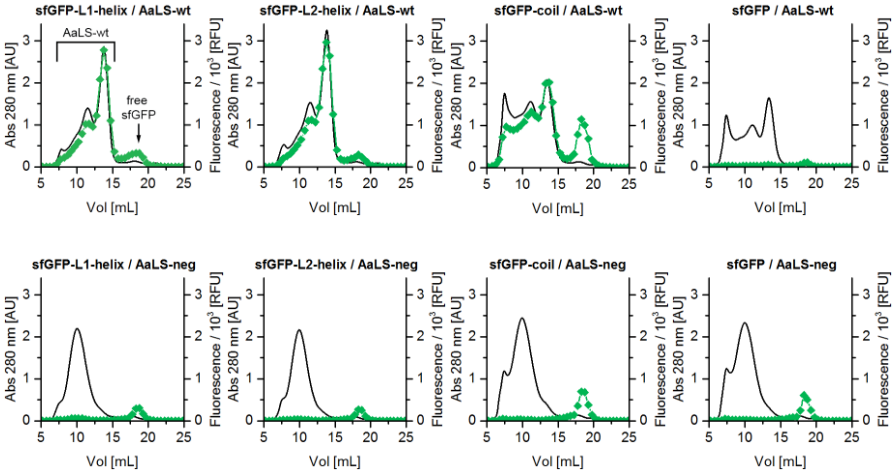


Figure 86. Chromatograms of sfGFP-tag/AaLS complexes. Black traces indicate on-line absorption at 280 nm. Green points indicate the fluorescence of individual fractions (excitation 488 nm, emission 508 nm).

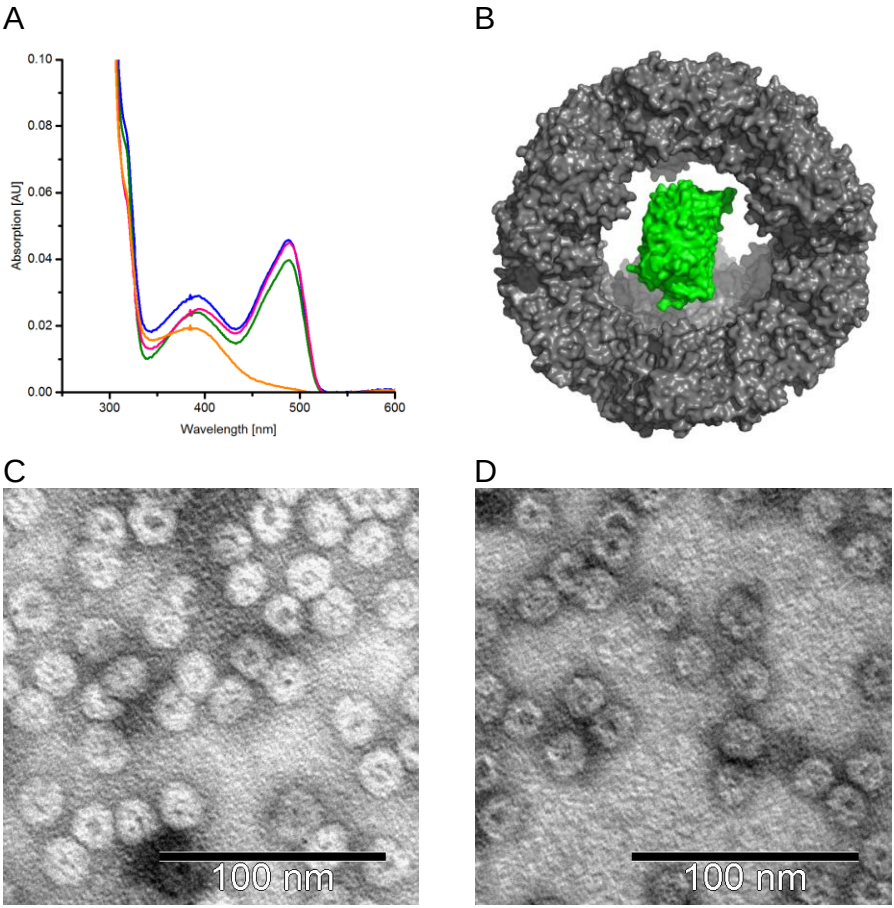


Figure 87. Characterization of sfGFP / AaLS-wt complexes. A) Absorption spectra of purified AaLS-wt capsids that were co-expressed with sfGFP-L1-helix (blue), sfGFP-L2-helix (green), sfGFP-coil (pink), and sfGFP (orange). The absorption maxima at 390 nm and 488 nm correspond to flavins bound in the AaLS active site, and sfGFP respectively. B) Rendering of a sfGFP molecule (green, PDB 2B3P) placed within an AaLS 60-mer (grey, PDB 1HQK). C) Electron microscopy image of sfGFP-coil / AaLS-wt. D) Image of empty AaLS-wt.

5.3 Discussion

Fusion to the native AaRS C-terminal peptide presents an interesting alternative to designed electrostatic interactions for targeting foreign proteins to the lumen of lumazine synthase capsids and offers access to a completely different capsid-cargo system. Whereas AaLS-neg and AaLS-13 can encapsulate up to 45 guests per 180-meric shell, and thus provide a high local protein density, AaLS-wt allows to isolate

a protein from its environment. The AaLS-neg and AaLS-13 shells enable easy access of macromolecules to its lumen and render the encapsulation of guest proteins reversible. The sfGFP/AaLS complexes have not yet been characterized with the same scrutiny, but are likely more stable and it will be interesting to see which conditions cause its disintegration.

Failure to form stable complexes between AaLS-neg and sfGFP decorated with the dodecameric peptide could have various causes. The peptide-protein interactions could have been weakened or abolished by conformational changes resulting from introduction of four negative charges [177]. This hypothesis is corroborated by solutions of AaLS-neg being less yellow than AaLS-wt, indicating that flavins do not bind as well to AaLS-neg as to AaLS-wt. Alternatively, the peptide-protein interaction might not have been strong enough to begin with for retaining guests in dilute solutions. AaLS-neg in comparison to AaLS-wt is likely more flexible as judged from swift encapsulation and ready release of supercharged protein guests (see Chapter 3). If the peptide-protein interaction is just strong enough to target guests to the lumen at the protein concentrations in the bacterial cytosol, they will be released under the dilute concentrations of the purification, unless the protein shell is impenetrable as might be the case for AaLS-wt, but not AaLS-neg. Protein binding could be potentially enhanced by increasing the valence of the interactions. Trimeric coiled-coils could be employed to oligomerize proteins of interest and present the C-terminal targeting peptides in an appropriate geometry [266–269]. If this approach were successful for encapsulating cargo proteins in AaLS-neg, one could utilize two orthogonal encapsulation strategies for creating more intricate host-guest complexes. Moreover, heterotrimeric coiled coils [270,271] could be employed to encapsulate different cargo proteins at defined stoichiometric ratios and close proximity.

Encapsulation of single protein within capsids is novel and is interesting for a variety of applications. It can be used for studying protein folding in isolation and thus disentangle intrinsic unimolecular effects and inter-molecular aggregation. The high local protein concentration within the capsid allows to examine the effect of crowding on protein properties, which might be particularly interesting for investigating intrinsically disordered enzymes [272]. Lastly, encapsulation in a thermostable capsid might stabilize fragile enzymes by protecting them from aggregation and proteolysis and thus facilitate their application in industrial biocatalysis.

5.4 Materials and methods

5.4.1 Molecular biology

pACTET_AaRS. A gene, codon-optimized for expression in *E. coli*, of AaRS was synthesized by ATG Biosynthetics (Germany) and provided on plasmid pGH_AaRS. This plasmid was digested with *Nde*I and *Spe*I, purified by agarose gel electrophoresis and the small fragment ligated into accordingly restricted plasmid pACTET_HIV-R10 [179] to give pACTET_AaRS.

pACYC_His6_sfGFP. The gene encoding sfGFP [264] was amplified from plasmid pQIq_off7_sfGFP, a generous gift from Prof. Andreas Plückthun [273], using primers BamHI-sfGFP_fw and sfGFP_rv. The resulting product was digested with *Bam*HI and *Spe*I and ligated into accordingly restricted and dephosphorylated pACYC_His6-GFP36p [227] to give pACYC_His6-GFP36p.

pACTET_sfGFP-L1-helix. The gene encoding sfGFP was amplified from plasmid pACYC_His6_sfGFP using primers NdeI-sfGFP_fw and L1_rv. The gene encoding the C-terminal helix of AaRS was amplified from pACTET_AaRS using primers L1_fw and AaRS-SpeI_rv. Both genes were assembled by overlap extension PCR using NdeI-sfGFP_fw and AaRS-SpeI_rv as flanking primers and the resulting gene digested with *Nde*I and *Spe*I before ligation into accordingly restricted plasmid pACTET_HIV-R10 to give pACTET_sfGFP-L1-helix. Plasmid pACTET_sfGFP-L2-helix was prepared analogously.

pACTET_sfGFP-coil. Plasmid pACTET_sfGFP-L1-helix was digested with *Bam*HI and *Spe*I. Oligonucleotides coil_fw and coil_rv were denatured for 5 min at 95 °C, and subsequently annealed by gradually cooling to 25 °C over 6 min. The resulting duplex DNA was ligated into pACTET_sfGFP-L1-helix that had been digested with *Bam*HI and *Spe*I to give pACTET_sfGFP-coil.

Table 11. Oligonucleotides used for molecular biology.

Name	Sequence
AaRS-SpeI_rv	CAGCTGACTAGTCACCATTTCAGAAACTCTTTGAAGATG
BamHI_sfGFP_fw	ACCACGGATCCGGTATGAGCAAAGGAGAAGAACTTTTC
coil_fw	GATCCGGTAAGAAAGAAGACATCTTCAAAGAGTTTCTGAAATGGTGA
coil_rv	CTAGTCACCATTTCAGAAACTCTTTGAAGATGTCTTCTTTCTTACCG
L1_fw	CAAAGGATCCGGTACGGACATCCTGGGCAAATACG
L1_rv	GGATGTCCGTACCGGATCCTTTGTAGAGTTCATCCATGC
L2_fw	CAAAGGATCCGGTGGTAGTGGTACGGACATCCTGGGCAAATACG
L2_rv	CCACTACCACCGGATCCTTTGTAGAGTTCATCCATGCC
NdeI-sfGFP_fw	GATATACATATGAGCAAAGGAGAAGAACTTTTC
sfGFP_rv	TCAGCTGACTAGTCATTTGTAGAGTTCATCCATGCC

5.4.2 Protein sequences

Aquifex aeolicus lumazine synthase (AaRS)

MFTGLVEDLGKVKNLTLSSKGAKLSVETKLEDVKLGDSVSVNGACLTVVDIKSST
LTFDVSPETLKRTNLGKLKTGDYVNLERALRVGERLGGHIVQGHVDFTAPVKSFN
FLGEHYELVIEIPEEWSIYVVEKGSIALDGISLTVNYVKENKVFINIIPHTYKST
NLQFKKVGDLLNVETDILGKYVINYLNKLKKKEDIFKEFLKW*

Superfolder green fluorescent protein (sfGFP)

MSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPV
PWPTLVTTLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGTYKTRA
EVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANF
KIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSVLSKDPNEKRDHMVL
LEFVTAAGITHGMDELYK*

sfGFP-L1-helix

MSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPV
PWPTLVTTLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGTYKTRA
EVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANF
KIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSVLSKDPNEKRDHMVL
LEFVTAAGITHGMDELYKGSGTDILGKYVINYLNKLKKKEDIFKEFLKW*

sfGFP-L2-helix

MSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPV
PWPTLVTTLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGTYKTRA
EVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANF

KIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSVLSKDPNEKRDHMVL
LEFVTAAGITHGMDELYKGSGGSGTDILGKYVINYLNKLKKKEDIFKEFLKW*

sfGFP-coil

MSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPV
PWPTLVTTLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGTYKTRA
EVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANF
KIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSVLSKDPNEKRDHMVL
LEFVTAAGITHGMDELYKGSGTDILGKYVINYLNKLKKKEDIFKEFLKW*

5.4.3 Protein production and purification

Plasmids pACTET in which transcription is placed under control of the tetracycline repressor encoding AaRS, sfGFP or tagged sfGFP variants were co-transformed with pMG211-AaLS* plasmids encoding different AaLS variants into *E. coli* KA13 cells [203]. Dense overnight cultures of these transformants were used to inoculate 400 mL LB (25 µg/mL chloramphenicol, 100 µg/mL ampicillin), which was then incubated at 37 °C, 230 rpm until OD_{600} ~0.6 at which point protein expression was induced by addition of 5 µg/mL tetracycline and 100 µM IPTG. The cell cultures were incubated at 30 °C, 230 rpm for 16 hours before the cell pellet was harvested by centrifugation and stored at -20 °C.

Cell pellets were thawed and suspended in lysis buffer (50 mM NaP_i, 300mM NaCl, pH 8.0), supplemented with protease inhibitor cocktail (cOmplete protease inhibitor cocktail tablets, Roche, Switzerland), spatula tips of DNase I, RNase A, 5 mM $MgCl_2$, 1 mg/mL lysozyme, 1 mg/mL polymixin and incubated for 1 h at room temperature prior to sonication. The cleared supernatant was loaded onto Ni-NTA agarose, washed with lysis buffer containing 20 mM and 40 mM imidazole, and finally eluted with lysis buffer containing 500 mM imidazole. Following elution, 5 mM EDTA was added. The eluted protein was concentrated and separated on a Superose 6 10/300 GL column (GE Healthcare, USA) at room temperature using 50 mM NaP_i, 200 mM NaCl, 5 mM EDTA, pH 8.0 as mobile phase. Fluorescence of individual fractions was measured on a plate reader.

6 Perspectives

Biotechnology is a proven tool for tackling humanity's grand challenges. Improved plants make it possible to grow food on barren soil, helping to feed an ever-growing population. Therapeutic proteins alleviate hitherto untreatable diseases, and engineered microorganisms present a sustainable alternative to the production of oil-based chemicals. Originally, living matter was employed for technical purposes in its unmodified form, but the advent of molecular biology only three decades ago has made it possible for researchers to alter the genome of entire organisms and thus tailor them towards specific tasks.

The emerging field of synthetic biology is attempting to go beyond incremental improvements of existing properties to create completely new biological entities. Entire metabolic pathways have been successfully installed in heterologous microorganisms, and artificial enzymes have been created from scratch. An important next step will address the assembly of multi-component higher-order structures that can execute more complex tasks, yet as independent synthetic biology building blocks. Artificial cellular organelles, comprising multiple enzymes performing a particular metabolic sequence and segregated by a shell, are one embodiment of this vision. Ideally, all components should be genetically encodable and assemble spontaneously in a cellular environment so that, once designed, they can be effortlessly replicated.

The overarching goal of this thesis was to use an engineered protein cage to construct multi-enzyme nanoreactors in vitro and conduct kinetic experiments on the resulting guest-host complex in order to identify the prerequisites for small-molecule channeling between the active sites of encapsulated enzymes. Toward this end, the mechanism of a computationally designed retro-aldolase RA95.5-8 was investigated. We identified carbinolamine formation or carbon-carbon bond cleavage as likely rate-limiting steps. Using stopped-flow spectroscopy and a narrow-bandwidth optical filter it should now be possible to determine individual rates of both steps. After the initial experiments characterizing RA95.5-8 were conducted, Obexer developed an ultra-high throughput microfluidic based screen for retro-aldolase activity and identified an improved variant RA95.5-8F with an approximately 60-fold increased catalytic proficiency as well as total turnover and reduced susceptibility to inhibition by the aldehyde product [153]. The experiments reported in Chapter 2 to characterize RA95.5-8 should be extended to RA95.5-8F to rationalize these effects. The results will not only guide future design efforts and further directed evolution experiments, but may also contribute to a general understanding of the mechanism used by class I aldolases.

The engineered capsid, AaLS-13, which was chosen as a host for the encapsulation experiments, was first characterized in detail with respect to kinetics of packaging and the stability of the resulting host-guest complexes. The FRET pair that was developed to measure the extent of encapsulation spectroscopically is likely to be useful for characterizing functional capsid–cargo complexes generated by many different encapsulation systems. In the case of AaLS-neg and AaLS-13, guest encapsulation was found to be unexpectedly fast, approaching the diffusion limit, and also completely reversible. Protein internalization appears to be driven entirely by the strength of the electrostatic interactions between guest and host. Remarkably, guest binding and release do not require disassembly of the capsid shell, indicating that the capsid shell either exhibits considerable conformational flexibility or possesses, if only transiently, large pores. Such phenomena have not been observed with virus-like particles or bacterial encapsulins, making AaLS-neg and AaLS-13 capsids uniquely valuable subjects for detailed investigations of capsid assembly and stability, and for applications that cannot be realized with previously described encapsulation systems [274]. The reversibility of guest encapsulation could be particularly valuable for delivery applications, for instance. Experiments are currently ongoing in our laboratory to exploit AaLS-neg and AaLS-13 to deliver biomacromolecules to living cells in therapeutic context. The ability to overproduce toxic proteins in *E. coli* strains by co-expression with AaLS-13 has been already demonstrated by Wörsdörfer [180]. Since the guests can be subsequently released from the capsids simply by increasing the ionic strength, this strategy could prove useful for the production of diverse toxins. Efforts to produce the bactericidal permeability increasing protein [275] recombinantly in *E. coli* as a fusion protein with GFP36+ by co-expression with AaLS-13 are currently under way.

The unexpected permeability of AaLS-neg and AaLS-13 shells is probably not a problem for substrate channeling. Douglas and co-workers have demonstrated that the shell of bacteriophage P22 capsids does not limit substrate diffusion either, but proteinaceous cargo cannot cross the wall of this capsid [104,105,238]. After all, if the capsid shell were to retain small molecules, it would likewise impede access of substrates to the active sites of the encapsulated enzymes. What seems to be the decisive factor for substrate channeling in multi-enzyme complexes is the likelihood that the substrate encounters the active site of the downstream enzyme and the probability of converting the bound substrate into product. The latter depends on the ratio between the k_{cat} of the rate-limiting step and the compounded rates for reversal of all previous steps leading up to the rate-limiting step, which – to a first approximation – should scale with k_{cat}/K_m of the downstream enzyme. The likeli-

hood of an intermediate encountering the active site of a downstream enzyme before escaping the realm of the complex depends on its size, the density with which enzymes are packed, and their spatial arrangement.

Calculations performed by Wingreen and co-workers suggest that, even for enzymes packed at high density, clusters need to have diameters greater than 100 nm to be effective channelers. Even the relatively large AaLS-13 capsid is more than one order of magnitude smaller than the optimal value of ca. 1 µm calculated. Additionally, cargo proteins are distributed along the inside of the protein shell rather than filling the lumenal space. Ideally, the upstream enzymes should be localized at the center of the capsid and surrounded by the downstream enzymes, so that the intermediate would have to pass them on their way out of the capsid. In principle, larger clusters of enzymes could be obtained by forming 3-dimensional assemblies of capsids [276,277]. In case of AaLS-neg or AaLS-13, the capsid exterior could be functionalized chemically or by genetic fusion to peptides that can form coiled-coils. Addition of a multivalent scaffold following guest loading could then effect formation of larger networks, the extent of which might be subject to the concentration at which the scaffold is added. Ironically, the tendency of AaLS-13 and AaLS-neg capsids to precipitate in the presence of excess guest, which has bedeviled many encapsulation attempts much to the chagrin of the experimentalists, might be successfully exploited to create enzyme clusters of the size required for effective metabolite channeling. Loading the capsid with a mixture of upstream and downstream enzymes, followed by addition of excess oligomeric downstream enzyme, would create aggregates in which the upstream enzymes are surrounded by a corona of downstream enzymes. If the capsids retain their integrity upon aggregation, the encapsulated enzymes should be protected and active. Effective metabolite channeling would then only depend on whether the substrates can penetrate the aggregate. If non-catalytic supercharged proteins or polymers were used to mediate association of multiple capsids, precise control over the stoichiometric ratios of encapsulated enzymes, which is a hallmark of the host-guest complexes formed by AaLS-neg and AaLS-13, could prevail and be further exploited in larger assemblies.

In summary, although the cascade reaction involving RA95.5-8 and horse liver alcohol dehydrogenase is not accelerated upon co-encapsulation in AaLS-neg, the options for improvement have not yet been exhausted. Owing to the extensive characterization of the AaLS capsids as well as the involved enzymes, their implementation should be relatively straightforward. The improved RA95.5-8F variant as well as alcohol and aldehyde dehydrogenases with higher activity towards naphthaldehydes should certainly be tested, since proficient enzymes are a prerequisite for substrate channeling. Assembly of higher order capsid structures of capsids

seems feasible and might allow systematic analysis of the influence of enzyme-domain size on channeling efficiency.

The various examples of natural multi-enzyme complexes, whether formed by strong protein-protein interactions, co-encapsulation within shells, or mediated transiently by membranes or the cytoskeleton, suggest that co-localizing biocatalysts in a metabolic pathway is evolutionarily advantageous. The precise mode by which such domains confer this advantage is currently a matter of passionate debate that would benefit from systematic in vitro kinetic experiments. The insights gained could further the development of synthetic metabolons and thus expedite possible biotechnological approaches for the production of chemicals.

Bibliography

1. Meadows DH, Club of Rome: **The limits to growth: A report for the Club of Rome's project on the predicament of mankind**. New York: Universe Books, 1972.
2. Nefiodow L, Nefiodow S: **The sixth Kondratieff: The new long wave in the global economy**. CreateSpace Independent Publishing Platform, 2014.
3. Michel H: **The nonsense of biofuels**. *Angew. Chemie - Int. Ed.* 2012, **51**:2516–2518.
4. Vasic-Racki D: **History of industrial biotransformations - Dreams and realities**. In *Industrial Biotransformations*. Edited by Liese A, Seelbach K, Wandrey C. Weinheim: Wiley-VCH, 2006: 1–35.
5. Neuberg C, Hirsch J: **Über ein Kohlenstoffketten knüpfendes Ferment (Carboligase)**. *Biochem. Z.* 1921, **115**:282–310.
6. Outtrup H, Norman BE: **Properties and application of a thermostable maltogenic amylase produced by a strain of Bacillus modified by recombinant-DNA techniques**. *Starch - Stärke* 1984, **36**:405–411.
7. Estell DA, Graycar TP, Wells JA: **Engineering an enzyme by site-directed mutagenesis to be resistant to chemical oxidation**. *J. Biol. Chem.* 1985, **260**:6518–6521.
8. Turner NJ: **Directed evolution drives the next generation of biocatalysts.** *Nat. Chem. Biol.* 2009, **5**:567–73.
9. Paddon CJ, Westfall PJ, Pitera DJ, Benjamin K, Fisher K, McPhee D, Leavell MD, Tai A, Main A, Eng D, et al.: **High-level semi-synthetic production of the potent antimalarial artemisinin**. *Nature* 2013, **496**:528–532.
10. Galanie S, Thodey K, Trenchard IJ, Filsinger Interrante M, Smolke CD: **Complete biosynthesis of opioids in yeast**. *Science* 2015, **349**:1095–1100.
11. Kries H, Blomberg R, Hilvert D: **De novo enzymes by computational design**. *Curr. Opin. Chem. Biol.* 2013, **17**:221–228.
12. Hilvert D: **Design of protein catalysts**. *Annu. Rev. Biochem.* 2013, **82**:447–470.
13. Lee H, DeLoache WC, Dueber JE: **Spatial organization of enzymes for metabolic engineering.** *Metab. Eng.* 2012, **14**:242–251.
14. DeLoache WC, Dueber JE: **Compartmentalizing metabolic pathways in organelles.** *Nat. Biotechnol.* 2013, **31**:320–1.
15. Srere PA: **The metabolon**. *Trends Biochem. Sci.* 1985, **10**:109–110.
16. Avalos JL, Fink GR, Stephanopoulos G: **Compartmentalization of metabolic pathways in yeast mitochondria improves the production of branched-chain alcohols**. *Nat. Biotechnol.* 2013, **31**:335–41.
17. Kim EY, Tullman-Ercek D: **Engineering nanoscale protein compartments for**

synthetic organelles. *Curr. Opin. Biotechnol.* 2013, **24**:627–32.
18. Fu J, Liu M, Liu Y, Woodbury NW, Yan H: **Interenzyme substrate diffusion for an enzyme cascade organized on spatially addressable DNA nanostructures**. *J. Am. Chem. Soc.* 2012, **134**:5516–5519.
19. Pathria RK, Beale PD: **Fluctuations and nonequilibrium statistical mechanics**. In *Statistical mechanics*. Edited by Pathria RK, Beale PD. Boston: Academic Press, 2011: 583–635.
20. Elowitz MB, Surette MG, Wolf PE, Stock JB, Leibler S: **Protein mobility in the cytoplasm of Escherichia coli**. *J. Bacteriol.* 1999, **181**:197–203.
21. Spivey HO, Ovádi J: **Substrate channeling**. *Methods* 1999, **19**:306–321.
22. Lin J, Palomec L, Wheeldon I: **Design and analysis of enhanced catalysis in scaffolded multienzyme cascade reactions**. *ACS Catal.* 2014, **4**:505–511.
23. Keleti T: **Channelling in enzyme complexes**. In *Dynamics of Biochemical Systems*. Edited by Ricard J, Cornish-Bowden A. Boston: Springer, 1984: 103–114.
24. Bryce CF, Williams DC, John RA, Fasella P: **The anomalous kinetics of coupled aspartate aminotransferase and malate dehydrogenase. Evidence for compartmentation of oxaloacetate.** *Biochem. J.* 1976, **153**:571–7.
25. Wolfram Research: **Mathematica**. Champaign (IL), USA: Wolfram Research, Inc., 2015.
26. Lin JL, Wheeldon I: **Kinetic enhancements in DNA-enzyme nanostructures mimic the Sabatier principle**. *ACS Catal.* 2013, **3**:560–564.
27. Elcock AH, Huber GA, Andrew McCammon J: **Electrostatic channeling of substrates between enzyme active sites: Comparison of simulation and experiment**. *Biochemistry* 1997, **36**:16049–16058.
28. Shatalin K, Lebreton S, Rault-Leonardon M, Vélot C, Srere PA: **Electrostatic channeling of oxaloacetate in a fusion protein of porcine citrate synthase and porcine mitochondrial malate dehydrogenase**. *Biochemistry* 1999, **38**:881–889.
29. Svedružić ŽM, Spivey HO: **Interaction between mammalian glyceraldehyde-3-phosphate dehydrogenase and L-lactate dehydrogenase from heart and muscle**. *Proteins Struct. Funct. Bioinforma.* 2006, **63**:501–511.
30. Srivastava DK, Smolen P, Betts GF, Fukushima T, Spivey HO, Bernhard SA: **Direct transfer of NADH between alpha-glycerol phosphate dehydrogenase and lactate dehydrogenase: fact or misinterpretation?** *Proc. Natl. Acad. Sci. U. S. A.* 1989, **86**:6464–6468.
31. Srivastava DK, Bernhard SA: **Direct transfer of reduced nicotinamide**

adenine dinucleotide from glyceraldehyde 3-phosphate dehydrogenase to liver alcohol dehydrogenase. *Biochemistry* 1984, **23**:4538–4545.
32. Srivastava DK, Bernhard SA: **Mechanism of transfer of reduced nicotinamide adenine dinucleotide among dehydrogenases.** *Biochemistry* 1985, **24**:623–8.
33. Wu XM, Gutfreund H, Lakatos S, Chock PB: **Substrate channeling in glycolysis: a phantom phenomenon.** *Proc. Natl. Acad. Sci. U. S. A.* 1991, **88**:497–501.
34. Patterson DP, Schwarz B, Waters RS, Gedeon T, Douglas T: **Encapsulation of an enzyme cascade within the bacteriophage P22 virus-like particle**. *ACS Chem. Biol.* 2014, **9**:359–365.
35. Castellana M, Wilson MZ, Xu Y, Joshi P, Cristea IM, Rabinowitz JD, Gitai Z, Wingreen NS: **Enzyme clustering accelerates processing of intermediates through metabolic channeling.** *Nat. Biotechnol.* 2014, **32**:1011–8.
36. Maier T, Leibundgut M, Ban N: **The crystal structure of a mammalian fatty acid synthase.** *Science* 2008, **321**:1315–22.
37. Izard T, Aevarsson A, Allen MD, Westphal AH, Perham RN, de Kok A, Hol WG: **Principles of quasi-equivalence and Euclidean geometry govern the assembly of cubic and dodecahedral cores of pyruvate dehydrogenase complexes.** *Proc. Natl. Acad. Sci. U. S. A.* 1999, **96**:1240–1245.
38. Patel MS, Nemeria NS, Furey W, Jordan F: **The pyruvate dehydrogenase complexes: Structure-based function and regulation**. *J. Biol. Chem.* 2014, **289**:16615–16623.
39. Kries H: **Tailor-made biocatalysts by enzyme design, redesign, and directed evolution**. Thesis 2014, ETH Zürich.
40. Huang X, Holden HM, Raushel FM: **Channeling of substrates and intermediates in enzyme-catalyzed reactions**. *Annu. Rev. Biochem.* 2001, **70**:149–180.
41. Miles EW: **Tryptophan synthase: A multienzyme complex with an intramolecular tunnel**. *Chem. Rec.* 2001, **1**:140–151.
42. Raboni S, Bettati S, Mozzarelli A: **Tryptophan synthase: A mine for enzymologists**. *Cell. Mol. Life Sci.* 2009, **66**:2391–2403.
43. Thoden JB, Holden HM, Wesenberg G, Raushel FM, Rayment I: **Structure of carbamoyl phosphate synthetase: A journey of 96 Å from substrate to product**. *Biochemistry* 1997, **36**:6305–6316.
44. Wang T, Bishop SH, Himoe A: **Detection of carbamate as a product of the carbamate kinase-catalyzed reaction by stopped flow spectrophotometry**. *J. Biol. Chem.* 1972, **247**:4437–4440.
45. Meek TD, Garvey EP, Santi D V: **Purification and characterization of the**

bifunctional thymidylate synthetase-dihydrofolate reductase from methotrexate-resistant Leishmania tropica. *Biochemistry* 1985, **24**:678–686.

46. Trujillo M, Donald RGK, Roos DS, Greene PJ, Santi D V: **Heterologous expression and characterization of the bifunctional dihydrofolate reductase-thymidylate synthase enzyme of Toxoplasma gondii**. *Biochemistry* 1996, **35**:6366–6374.
47. Liang PH, Anderson KS: **Substrate channeling and domain-domain interactions in bifunctional thymidylate synthase-dihydrofolate reductase**. *Biochemistry* 1998, **37**:12195–12205.
48. Sharma H, Landau MJ, Vargo MA, Spasov KA, Anderson KS: **First three-dimensional structure of toxoplasma gondii thymidylate synthase-dihydrofolate reductase: Insights for catalysis, interdomain interactions, and substrate channeling**. *Biochemistry* 2013, **52**:7305–7317.
49. Knighton DR, Kan C-C, Howland E, Janson CA, Hostomska Z, Welsh KM, Matthews DA: **Structure of and kinetic channelling in bifunctional dihydrofolate reductase–thymidylate synthase**. *Nat. Struct. Biol.* 1994, **1**:186–194.
50. Wang N, McCammon JA: **Substrate channeling between the human dihydrofolate reductase and thymidylate synthase**. *Protein Sci.* 2016, **25**:79–86.
51. Stroud RM: **An electrostatic highway**. *Nat. Struct. Biol.* 1994, **1**:131–134.
52. Yuvaniyama J, Chitnumsub P, Kamchonwongpaisan S, Vanichtanankul J, Sirawaraporn W, Taylor P, Walkinshaw MD, Yuthavong Y: **Insights into antifolate resistance from malarial DHFR-TS structures**. *Nat. Struct. Biol.* 2003, **10**:357–365.
53. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA: **Electrostatics of nanosystems: Application to microtubules and the ribosome.** *PNAS* 2001, **98**:10037–10041.
54. Schrödinger L: *The PyMOL Molecular Graphics System, Version 1.7.4*.
55. Srere PA: **Complexes of sequential metabolic enzymes.** *Annu. Rev. Biochem.* 1987, **56**:89–124.
56. An S, Deng Y, Tomsho JW, Kyoung M, Benkovic SJ: **Microtubule-assisted mechanism for functional metabolic macromolecular complex formation.** *Proc. Natl. Acad. Sci. U. S. A.* 2010, **107**:12872–12876.
57. Winkel BSJ: **Metabolic channeling in plants**. *Annu. Rev. Plant Biol.* 2004, **55**:85–107.
58. Jørgensen K, Rasmussen AV, Morant M, Nielsen AH, Bjarnholt N, Zagrobelny M, Bak S, Møller BL: **Metabolon formation and metabolic**

channeling in the biosynthesis of plant natural products. *Curr. Opin. Plant Biol.* 2005, **8**:280–291.
59. Laursen T, Møller BL, Bassard J-E: **Plasticity of specialized metabolism as mediated by dynamic metabolons**. *Trends Plant Sci.* 2015, **20**:20–32.
60. Suss KH, Arkona C, Manteuffel R, Adler K: **Calvin cycle multienzyme complexes are bound to chloroplast thylakoid membranes of higher plants in situ.** *Proc. Natl. Acad. Sci. U. S. A.* 1993, **90**:5514–8.
61. Stafford HA: **Possible multi-enzyme complexes regulating the formation of C6-C3 phenolic compounds and lignins in higher plants**. *Rec. Adv. Phytochem.* 1974, **8**:53–79.
62. Welch GR, Gaertner FH: **Enzyme organization in the polyaromatic-biosynthetic pathway: The arom conjugate and other multienzyme systems**. *Curr. Top. Cell. Regul.* 1980, **16**:113–162.
63. Lampkin SLI V, Cole KW, Vitto A, Gaertner FH: **The protease problem in Neurospora**. *Arch. Biochem. Biophys.* 1976, **177**:561–565.
64. Vitro A, Gaertner FH: **Proteolytic inactivation of a pentafunctional enzyme conjugate: Coordinate protection by the first substrate**. *Biochem. Biophys. Res. Commun.* 1978, **82**:977–981.
65. Welch GR, Gaertner FH: **Coordinate activation of a multienzyme complex by the first substrate**. *Arch. Biochem. Biophys.* 1976, **172**:476–489.
66. Rasmussen S, Dixon R: **Transgene-mediated and elicitor-induced perturbation of metabolic channeling at the entry point into the phenylpropanoid pathway**. *Plant Cell* 1999, **11**:1537–1552.
67. Achnine L, Blancaflor EB, Rasmussen S, Dixon RA: **Colocalization of L-phenylalanine ammonia-lyase and cinnamate 4-hydroxylase for metabolic channeling in phenylpropanoid biosynthesis**. *Plant Cell* 2004, **16**:3098–3109.
68. Borejsza-Wysocki W, Hrazdina G: **Aromatic polyketide synthases. Purification, characterization, and antibody development to benzalacetone synthase from raspberry fruits**. *Plant Physiol.* 1996, **110**:791–799.
69. Austin MB, Noel JP: **The chalcone synthase superfamily of type III polyketide synthases**. *Nat. Prod. Rep.* 2003, **20**:79–110.
70. Borejsza-Wysocki W, Hrazdina G: **Biosynthesis of p-hydroxyphenylbutan-2-one in raspberry fruits and tissue cultures**. *Phytochemistry* 1993, **35**:623–628.
71. Gleadow RM, Woodrow IE: **Constraints on effectiveness of cyanogenic glycosides in herbivore defense**. *J. Chem. Ecol.* 2002, **28**:1301–1313.
72. Møller BL, Conn EE: **The biosynthesis of cyanogenic glucosides in higher**

plants. *J. Biol. Chem.* 1979, **254**:8575–8583.

73. Kahn RA, Fahrendorf T, Halkier BA, Møller BL: **Substrate specificity of the cytochrome P450 enzymes CYP79A1 and CYP71E1 involved in the biosynthesis of the cyanogenic glucoside dhurrin in Sorghum bicolor (L.) Moench.** *Arch. Biochem. Biophys.* 1999, **363**:9–18.
74. Nielsen KA, Tattersall DB, Jones PR, Møller BL: **Metabolon formation in dhurrin biosynthesis**. *Phytochemistry* 2008, **69**:88–98.
75. Kristensen C, Morant M, Olsen CE, Ekstrøm CT, Galbraith DW, Møller BL, Bak S: **Metabolic engineering of dhurrin in transgenic Arabidopsis plants with marginal inadvertent effects on the metabolome and transcriptome.** *Proc. Natl. Acad. Sci. U. S. A.* 2005, **102**:1779–1784.
76. An S, Kumar R, Sheets ED, Benkovic SJ: **Reversible compartmentalization of de novo purine biosynthetic complexes in living cells.** *Science* 2008, **320**:103–106.
77. Zhao A, Tsechansky M, Swaminathan J, Cook L, Ellington AD, Marcotte EM: **Transiently transfected purine biosynthetic enzymes form stress bodies.** *PLoS One* 2013, **8**:e56203.
78. Knull HR, Walsh JL: **Association of Glycolytic Enzymes with the Cytoskeleton**. *Curr. Top. Cell. Regul.* 1992, **33**:15–30.
79. Menard L, Maughan D, Vigoreaux J: **The structural and functional coordination of glycolytic enzymes in muscle: Evidence of a Metabolon?** *Biology (Basel).* 2014, **3**:623–644.
80. Campanella ME, Chu H, Low PS: **Assembly and regulation of a glycolytic enzyme complex on the human erythrocyte membrane.** *Proc. Natl. Acad. Sci. U. S. A.* 2005, **102**:2402–2407.
81. Puchulu-Campanella E, Chu H, Anstee DJ, Galan JA, Tao WA, Low PS: **Identification of the components of a glycolytic enzyme metabolon on the human red blood cell membrane**. *J. Biol. Chem.* 2013, **288**:848–858.
82. Chu H, Puchulu-Campanella E, Galan JA, Tao WA, Low PS, Hoffman JF: **Identification of cytoskeletal elements enclosing the ATP pools that fuel human red blood cell membrane cation pumps**. *Proc. Natl. Acad. Sci.* 2012, **109**:12794–12799.
83. Discher BM, Won YY, Ege DS, Lee JCM, Bates FS, Discher DE, Hammer DA: **Polymersomes: Tough vesicles made from diblock copolymers**. *Science* 1999, **284**:1143–1146.
84. Du J, O'Reilly RK: **Advances and challenges in smart and functional polymer vesicles**. *Soft Matter* 2009, **5**:3544.
85. Delcea M, Möhwald H, Skirtach AG: **Stimuli-responsive LbL capsules and nanoshells for drug delivery**. *Adv. Drug Deliv. Rev.* 2011, **63**:730–747.

86. Marguet M, Bonduelle C, Lecommandoux S: **Multicompartmentalized polymeric systems: towards biomimetic cellular structure and function**. *Chem. Soc. Rev.* 2013, **42**:512–529.
87. Peters RJRW, Louzao I, van Hest JCM: **From polymeric nanoreactors to artificial organelles**. *Chem. Sci.* 2012, **3**:335–342.
88. Lu A, O'Reilly RK: **Advances in nanoreactor technology using polymeric nanostructures**. *Curr. Opin. Biotechnol.* 2013, **24**:639–645.
89. Vriezema DM, Garcia PML, Sancho Oltra N, Hatzakis NS, Kuiper SM, Nolte RJM, Rowan AE, van Hest JCM: **Positional assembly of enzymes in polymersome nanoreactors for cascade reactions**. *Angew. Chemie Int. Ed.* 2007, **46**:7378–7382.
90. Kuiper SM, Nallani M, Vriezema DM, Cornelissen JJLM, van Hest JCM, Nolte RJM, Rowan AE: **Enzymes containing porous polymersomes as nano reaction vessels for cascade reactions.** *Org. Biomol. Chem.* 2008, **6**:4315–4318.
91. Tanner P, Onaca O, Balasubramanian V, Meier W, Palivan CG: **Enzymatic cascade reactions inside polymeric nanocontainers: A means to combat oxidative stress**. *Chem. - A Eur. J.* 2011, **17**:4552–4560.
92. Meeuwissen SA, Rioz-Martínez A, de Gonzalo G, Fraaije MW, Gotor V, van Hest JCM: **Cofactor regeneration in polymersome nanoreactors: enzymatically catalysed Baeyer–Villiger reactions**. *J. Mater. Chem.* 2011, **21**:18923.
93. Peters RJRW, Marguet M, Marais S, Fraaije MW, van Hest JCM, Lecommandoux S: **Cascade reactions in multicompartmentalized polymersomes**. *Angew. Chemie Int. Ed.* 2014, **53**:146–150.
94. Hosta-Rigau L, York-Duran MJ, Zhang Y, Goldie KN, Städler B: **Confined multiple enzymatic (cascade) reactions within poly(dopamine)-based capsosomes.** *ACS Appl. Mater. Interfaces* 2014, **6**:12771–9.
95. Huang X, Li M, Mann S: **Membrane-mediated cascade reactions by enzyme–polymer proteinosomes**. *Chem. Commun.* 2014, **50**:6278.
96. Huang X, Li M, Green DC, Williams DS, Patil AJ, Mann S: **Interfacial assembly of protein-polymer nano-conjugates into stimulus-responsive biomimetic protocells.** *Nat. Commun.* 2013, **4**:2239.
97. Rocha-Martín J, Rivas B de Las, Muñoz R, Guisán JM, López-Gallego F: **Rational co-immobilization of bi-enzyme cascades on porous supports and their applications in bio-redox reactions with insitu recycling of soluble cofactors**. *ChemCatChem* 2012, **4**:1279–1288.
98. Rocha-Martin J, Acosta A, Guisan JM, López-Gallego F: **Immobilizing systems biocatalysis for the selective oxidation of glycerol coupled to in**

situ cofactor recycling and hydrogen peroxide elimination. *ChemCatChem* 2015, **7**:1939–1947.

99. Zore O V, Pattammattel A, Gnanaguru S, Kumar C V, Kasi RM: **Bienzyme-polymer-graphene oxide quaternary hybrid biocatalysts : Efficient substrate channeling under chemically and thermally denaturing conditions**. *ACS Catal.* 2015, **5**:4979–4988.
100. Yang YR, Liu Y, Yan H: **DNA nanostructures as programmable biomolecular scaffolds**. *Bioconjug. Chem.* 2015, **26**:1381–1395.
101. Roberts CC, Chang CA: **Modeling of enhanced catalysis in multienzyme nanostructures: Effect of molecular scaffolds, spatial organization, and concentration**. *J. Chem. Theory Comput.* 2015, **11**:286–292.
102. Fu J, Yang YR, Johnson-Buck A, Liu M, Liu Y, Walter NG, Woodbury NW, Yan H: **Multi-enzyme complexes on DNA scaffolds capable of substrate channelling with an artificial swinging arm**. *Nat. Nanotechnol.* 2014, **9**:531–536.
103. Yeates TO, Crowley CS, Tanaka S: **Bacterial microcompartment organelles: protein shell structure and evolution.** *Annu. Rev. Biophys.* 2010, **39**:185–205.
104. Patterson DP, Schwarz B, El-Boubbou K, van der Oost J, Prevelige PE, Douglas T: **Virus-like particle nanoreactors: Programmed encapsulation of the thermostable CelB glycosidase inside the P22 capsid**. *Soft Matter* 2012, **8**:10158–66.
105. Patterson DP, Prevelige PE, Douglas T: **Nanoreactors by programmed enzyme encapsulation inside the capsid of the bacteriophage P22**. *ACS Nano* 2012, **6**:5000–9.
106. Conrado RJ, Varner JD, DeLisa MP: **Engineering the spatial organization of metabolic enzymes: mimicking nature's synergy**. *Curr. Opin. Biotechnol.* 2008, **19**:492–499.
107. Agapakis CM, Boyle PM, Silver PA: **Natural strategies for the spatial optimization of metabolism in synthetic biology.** *Nat. Chem. Biol.* 2012, **8**:527–35.
108. Chen AH, Silver PA: **Designing biological compartmentalization**. *Trends Cell Biol.* 2012, **22**:662–670.
109. Idan O, Hess H: **Engineering enzymatic cascades on nanoscale scaffolds**. *Curr. Opin. Biotechnol.* 2013, **24**:606–611.
110. Chen R, Chen Q, Kim H, Siu K-H, Sun Q, Tsai S-L, Chen W: **Biomolecular scaffolds for enhanced signaling and catalytic efficiency**. *Curr. Opin. Biotechnol.* 2014, **28**:59–68.
111. Siu K, Chen RP, Sun Q, Chen L, Tsai S, Chen W: **Synthetic scaffolds for**

pathway enhancement. *Curr. Opin. Biotechnol.* 2015, **36**:98–106.
112. Dueber JE, Wu GC, Malmirchegini GR, Moon TS, Petzold CJ, Ullal A V, Prather KLJ, Keasling JD: **Synthetic protein scaffolds provide modular control over metabolic flux.** *Nat. Biotechnol.* 2009, **27**:753–759.
113. Delebecque CJ, Lindner AB, Silver PA, Aldaye FA: **Organization of intracellular reactions with rationally designed RNA assemblies.** *Science* 2011, **333**:470–4.
114. Conrado RJ, Wu GC, Boock JT, Xu H, Chen SY, Lebar T, Turnsek J, Tomsic N, Avbelj M, Gaber R, et al.: **DNA-guided assembly of biosynthetic pathways promotes improved catalytic efficiency**. *Nucleic Acids Res.* 2012, **40**:1879–1889.
115. Woolston BM, Edgar S, Stephanopoulos G: **Metabolic engineering: Past and future**. *Annu. Rev. Chem. Biomol. Eng.* 2013, **4**:259–88.
116. Moon TS, Dueber JE, Shiue E, Prather KLJ: **Use of modular, synthetic scaffolds for improved production of glucaric acid in engineered E. coli.** *Metab. Eng.* 2010, **12**:298–305.
117. Agapakis CM, Ducat DC, Boyle PM, Wintermute EH, Way JC, Silver PA: **Insulation of a synthetic hydrogen metabolism circuit in bacteria.** *J. Biol. Eng.* 2010, **4**:3.
118. Sachdeva G, Garg A, Godding D, Way JC, Silver PA: **In vivo co-localization of enzymes on RNA scaffolds increases metabolic production in a geometrically dependent manner**. *Nucleic Acids Res.* 2014, **42**:9493–9503.
119. Wu S, Schalk M, Clark A, Miles RB, Coates R, Chappell J: **Redirection of cytosolic or plastidic isoprenoid precursors elevates terpene production in plants**. *Nat. Biotechnol.* 2006, **24**:1441–1447.
120. Filice M, Palomo JM: **Cascade reactions catalyzed by bionanostructures**. *ACS Catal.* 2014, **4**:1588–1598.
121. Yu C, Zhao H, Pugh RJ, Pedley AM, French J, Jones SA: **Purinosome formation as a function of the cell cycle**. *Proc. Natl. Acad. Sci.* 2014, **112**:1368–1373.
122. Minten IJ, Claessen VI, Blank K, Rowan AE, Nolte RJM, Cornelissen JJLM: **Catalytic capsids: the art of confinement**. *Chem. Sci.* 2011, **2**:358.
123. Zhang Y, Ge J, Liu Z: **Enhanced activity of immobilized or chemically modified enzymes**. *ACS Catal.* 2015, **5**:4503–4513.
124. Ramachandran S, Fontanille P, Pandey A, Larroche C: **Gluconic acid: Properties, applications and microbial production**. *Food Technol. Biotechnol.* 2006, **44**:185–195.
125. Ogawa K, Nakajima-Kambe T, Nakahara T, Kokufuta E: **Coimmobilization of gluconolactonase with glucose oxidase for improvement in kinetic**

property of enzymatically induced volume collapse in ionic gels. *Biomacromolecules* 2002, **3**:625–631.

126. Fesko K, Gruber-Khadjawi M: **Biocatalytic methods for C-C bond formation**. *ChemCatChem* 2013, **5**:1248–1272.
127. Windle CL, Müller M, Nelson A, Berry A: **Engineering aldolases as biocatalysts**. *Curr. Opin. Chem. Biol.* 2014, **19**:25–33.
128. Du J, Say RF, Lü W, Fuchs G, Einsle O: **Active-site remodelling in the bifunctional fructose-1,6-bisphosphate aldolase/phosphatase**. *Nature* 2011, **478**:534–537.
129. Cooper SJ, Leonard GA, McSweeney SM, Thompson AW, Naismith JH, Qamar S, Plater A, Berry A, Hunter WN: **The crystal structure of a class II fructose-1,6-bisphosphate aldolase shows a novel binuclear metal-binding active site embedded in a familiar fold**. *Structure* 1996, **4**:1303–1315.
130. Schürmann M, Sprenger GA: **Fructose-6-phosphate aldolase is a novel class I aldolase from Escherichia coli and is related to a novel group of bacterial transaldolases.** *J. Biol. Chem.* 2001, **276**:11055–61.
131. Rale M, Schneider S, Sprenger GA, Samland AK, Fessner WD: **Broadening deoxysugar glycodiversity: Natural and engineered transaldolases unlock a complementary substrate space**. *Chem. - A Eur. J.* 2011, **17**:2623–2632.
132. Schneider S, Gutiérrez M, Sandalova T, Schneider G, Clapés P, Sprenger GA, Samland AK: **Redesigning the Active Site of Transaldolase TalB from Escherichia coli: New Variants with Improved Affinity towards Nonphosphorylated Substrates**. *ChemBioChem* 2010, **11**:681–690.
133. Garrabou X, Joglar J, Parella T, Crehuet R, Bujons J, Clapés P: **Redesign of the phosphate binding site of L-rhamnulose-1-phosphate aldolase towards a dihydroxyacetone dependent aldolase**. *Adv. Synth. Catal.* 2011, **353**:89–99.
134. Jiang L, Althoff EA, Clemente FR, Doyle L, Röthlisberger D, Zanghellini A, Gallaher JL, Betker JL, Tanaka F, Barbas CF, et al.: **De novo computational design of retro-aldol enzymes.** *Science* 2008, **319**:1387–91.
135. Tantillo DJ, Chen J, Houk KN: **Theozymes and compuzymes: theoretical models for biological catalysis.** *Curr. Opin. Chem. Biol.* 1998, **2**:743–750.
136. Althoff EA, Wang L, Jiang L, Giger L, Lassila JK, Wang Z, Smith M, Hari S, Kast P, Herschlag D, et al.: **Robust design and optimization of retroaldol enzymes.** *Protein Sci.* 2012, **21**:717–26.
137. Giger L, Caner S, Obexer R, Kast P, Baker D, Ban N, Hilvert D: **Evolution of a designed retro-aldolase leads to complete active site remodeling**. *Nat. Chem. Biol.* 2013, **9**:494–8.

138. Giger L: **Directed evolution as a versatile tool to investigate natural and artificial aldolases**. Thesis 2011, ETH Zürich.
139. Bürgi HB, Dunitz JD, Lehn JM, Wipff G: **Stereochemistry of reaction paths at carbonyl centres**. *Tetrahedron* 1974, **30**:1563–1572.
140. Simón L, Goodman JM: **Enzyme catalysis by hydrogen bonds: the balance between transition state binding and substrate binding in oxyanion holes.** *J. Org. Chem.* 2010, **75**:1831–40.
141. Jencks WP: **Studies on the mechanism of oxime and semicarbazone formation**. *J. Am. Chem. Soc.* 1959, **81**:475–481.
142. Greenzaid P, Luz Z, Samuel D: **N.M.R. study of reversible hydration of aliphatic aldehydes and ketones. Isotopic oxygen exchange of acetone**. *Trans. Faraday Soc.* 1968, **64**:2780–2786.
143. Shulman A, Sitry D, Shulman H, Keinan E: **Highly efficient antibody-catalyzed deuteration of carbonyl compounds**. *Chem. - A Eur. J.* 2002, **8**:229–239.
144. Layer RW: **The chemistry of imines**. *Chem. Rev.* 1966, **63**:489–510.
145. Krężel A, Bal W: **A formula for correlating pKa values determined in D2O and H2O**. *J. Inorg. Biochem.* 2004, **98**:161–166.
146. Cioni P, Strambini GB: **Effect of heavy water on protein flexibility.** *Biophys. J.* 2002, **82**:3246–3253.
147. Garrabou X, Beck T, Hilvert D: **A promiscuous de novo retro-aldolase catalyzes asymmetric Michael additions via Schiff base intermediates**. *Angew. Chemie Int. Ed.* 2015, **54**:5609–5612.
148. Blomberg R, Kries H, Pinkas DM, Mittl PRE, Grütter MG, Privett HK, Mayo SL, Hilvert D: **Precision is essential for efficient catalysis in an evolved Kemp eliminase.** *Nature* 2013, **503**:418–21.
149. Khare SD, Kipnis Y, Greisen PJ, Takeuchi R, Ashani Y, Goldsmith M, Song Y, Gallaher JL, Silman I, Leader H, et al.: **Computational redesign of a mononuclear zinc metalloenzyme for organophosphate hydrolysis**. *Nat. Chem. Biol.* 2012, **8**:294–300.
150. Preiswerk N, Beck T, Schulz JD, Milovnik P, Mayer C, Siegel JB, Baker D, Hilvert D: **Impact of scaffold rigidity on the design and evolution of an artificial Diels-Alderase**. *Proc. Natl. Acad. Sci.* 2014, **111**:8013–8018.
151. Khersonsky O, Kiss G, Rothlisberger D, Dym O, Albeck S, Houk KN, Baker D, Tawfik DS: **Bridging the gaps in design methodologies by evolutionary optimization of the stability and proficiency of designed Kemp eliminase KE59**. *Proc. Natl. Acad. Sci.* 2012, **109**:10358–10363.
152. Obexer R, Studer S, Giger L, Pinkas DM, Grütter MG, Baker D, Hilvert D: **Active site plasticity of a computationally designed retro-aldolase**

enzyme. *ChemCatChem* 2014, **6**:1043–1050.
153. Obexer RJ: **Exploring the evolvability of artificial aldolases**. Thesis 2015, ETH Zürich.
154. Covington AK, Paabo M, Robinson RA, Bates RG: **Use of the glass electrode in deuterium oxide and the relation between the standardized pD (paD) scale and the operational pH in heavy water**. *Anal. Chem.* 1968, **40**:700–706.
155. Lwoff A, Anderson T, Jacob F: **Remarques sur les caractéristiques de la particule virale infectieuse**. *Ann. Inst. Pasteur (Paris).* 1959, **97**:281–289.
156. Caspar DLD, Dulbecco R, Klug A, Lwoff A, Stoker MGP, Tournier P, Wildy P: **Proposals**. *Cold Spring Harb. Symp. Quant. Biol.* 1962, **27**:49–50.
157. Caspar DLD, Klug A: **Physical principles in the construction of regular viruses**. *Cold Spring Harb. Symp. Quant. Biol.* 1962, **27**:1–24.
158. Darwin C: **On the Origin of Species**. London: John Murray, 1859.
159. Zhang X, Meining W, Fischer M, Bacher A, Ladenstein R: **X-ray structure analysis and crystallographic refinement of lumazine synthase from the hyperthermophile Aquifex aeolicus at 1.6 A resolution: determinants of thermostability revealed from structural comparisons.** *J. Mol. Biol.* 2001, **306**:1099–1114.
160. Pauling L: **Protein interactions. Aggregation of globular proteins**. *Discuss. Faraday Soc.* 1953, **13**:170.
161. Johnson JE, Speir JA: **Quasi-equivalent viruses: a paradigm for protein assemblies.** *J. Mol. Biol.* 1997, **269**:665–75.
162. Tanaka S, Kerfeld CA, Sawaya MR, Cai F, Heinhorst S, Cannon GC, Yeates TO: **Atomic-level models of the bacterial carboxysome shell.** *Science* 2008, **319**:1083–1086.
163. Klug A: **Structure of viruses of the papilloma-polyoma type**. *J. Mol. Biol.* 1965, **11**:424–431.
164. Prevelige PE, Thomas D, King J: **Nucleation and growth phases in the polymerization of coat and scaffolding subunits into icosahedral procapsid shells**. *Biophys. J.* 1993, **64**:824–835.
165. Bahadur RP, Rodier F, Janin J: **A dissection of the protein–protein interfaces in icosahedral virus capsids**. *J. Mol. Biol.* 2007, **367**:574–590.
166. Ceres P, Zlotnick A: **Weak protein-protein interactions are sufficient to drive assembly of Hepatitis B virus capsids**. *Biochemistry* 2002, **4**:11525–11531.
167. Katen S, Zlotnick A: **The thermodynamics of virus capsid assembly**. *Methods Enzymol.* 2009, **455**:395–417.
168. Zlotnick A, Fane BA: **Mechanisms of Icosahedral Virus Assembly**. *Struct.*

Virol. 2011, doi:10.1039/9781849732239-00180.
169. Perlmutter JD, Hagan MF: **Mechanisms of virus assembly**. *Annu. Rev. Phys. Chem.* 2015, **66**:217–239.
170. Šiber A, Božič AL, Podgornik R: **Energies and pressures in viruses: contribution of nonspecific electrostatic interactions**. *Phys. Chem. Chem. Phys.* 2011, **14**:3746–3765.
171. Singh S, Zlotnick A: **Observed hysteresis of virus capsid disassembly is implicit in kinetic models of assembly.** *J. Biol. Chem.* 2003, **278**:18249–18255.
172. Yeates TO, Kerfeld CA, Heinhorst S, Cannon GC, Shively JM: **Protein-based organelles in bacteria: carboxysomes and related microcompartments**. *Nat. Rev. Microbiol.* 2008, **6**:681–691.
173. Sutter M, Boehringer D, Gutmann S, Günther S, Prangishvili D, Loessner MJ, Stetter KO, Weber-Ban E, Ban N: **Structural basis of enzyme encapsulation into a bacterial nanocompartment.** *Nat. Struct. Mol. Biol.* 2008, **15**:939–47.
174. Ladenstein R, Fischer M, Bacher A: **The lumazine synthase/riboflavin synthase complex: shapes and functions of a highly variable enzyme system**. *FEBS J.* 2013, **280**:2537–2563.
175. Kis K, Bacher A: **Substrate channeling in the lumazine synthase/riboflavin synthase complex of Bacillus subtilis**. *J. Biol. Chem.* 1995, **270**:16788–16795.
176. Kjer-Nielsen L, Patel O, Corbett AJ, Le Nours J, Meehan B, Liu L, Bhati M, Chen Z, Kostenko L, Reantragoon R, et al.: **MR1 presents microbial vitamin B metabolites to MAIT cells**. *Nature* 2012, **491**:717–725.
177. Seebeck FP, Woycechowsky KJ, Zhuang W, Rabe JP, Hilvert D: **A simple tagging system for protein encapsulation**. *JACS* 2006, **128**:4516–17.
178. Seebeck FP: **Active sites at protein interfaces**. Thesis 2005, ETH Zürich.
179. Wörsdörfer B, Woycechowsky KJ, Hilvert D: **Directed evolution of a protein container**. *Science* 2011, **331**:589–592.
180. Wörsdörfer B: **Functionalizing capsid forming proteins**. Thesis 2010, ETH Zürich.
181. Wörsdörfer B, Pianowski Z, Hilvert D: **Efficient in vitro encapsulation of protein cargo by an engineered protein container**. *JACS* 2012, **134**:909–911.
182. Fuchs SM, Raines RT: **Arginine grafting to endow cell permeability**. *ACS Chem. Biol.* 2007, **2**:167–70.
183. Lawrence MS, Phillips KJ, Liu DR: **Supercharging proteins can impart unusual resilience.** *J. Am. Chem. Soc.* 2007, **129**:10110–10112.

184. Beck T, Tetter S, Künzle M, Hilvert D: **Construction of matryoshka-type structures from supercharged protein nanocages**. *Angew. Chemie Int. Ed.* 2015, **54**:937–940.
185. O'Neil A, Prevelige PE, Basu G, Douglas T: **Coconfinement of fluorescent proteins: Spatially enforced communication of GFP and mCherry encapsulated within the P22 capsid**. *Biomacromolecules* 2012, **13**:3902–3907.
186. Rurup WF, Verbij F, Koay MST, Blum C, Subramaniam V, Cornelissen JJLM: **Predicting the loading of virus-like particles with fluorescent proteins**. *Biomacromolecules* 2014, **15**:558–563.
187. Cubitt AB, Woollenweber LA, Heim R: **Understanding structure-function relationships in the Aequorea victoria green fluorescent protein.** *Methods Cell Biol.* 1999, **58**:19–30.
188. Wachter RM, Remington SJ: **Sensitivity of the yellow variant of green fluorescent protein to halides and nitrate**. *Curr. Biol.* 1999, **9**:R628–R629.
189. Patterson GH, Piston DW, Barisas BG: **Förster distances between green fluorescent protein pairs.** *Anal. Biochem.* 2000, **284**:438–440.
190. Fábián ÁI, Rente T, SzölloSi J, Matyus L, Jenei A: **Strength in numbers: Effects of acceptor abundance on FRET efficiency**. *ChemPhysChem* 2010, **11**:3713–3721.
191. Zheng J: **Biomedical Applications of Biophysics**. In *Handbook of Modern Biophysics*. Edited by Jue T. Totowa, NJ: Humana Press, 2010: 119–136.
192. Bai X, McMullan G, Scheres SH.: **How cryo-EM is revolutionizing structural biology**. *Trends Biochem. Sci.* 2015, **40**:49–57.
193. Snijder J, Van De Waterbeemd M, Damoc E, Denisov E, Grinfeld D, Bennett A, Agbandje-Mckenna M, Makarov A, Heck AJR: **Defining the stoichiometry and cargo load of viral and bacterial nanoparticles by orbitrap mass spectrometry**. *J. Am. Chem. Soc.* 2014, **136**:7295–7299.
194. Stano P, D'Aguanno E, Bolz J, Fahr A, Luisi PL: **A remarkable self-organization process as the origin of primitive functional cells**. *Angew. Chemie - Int. Ed.* 2013, **52**:13397–13400.
195. Stano P, de Souza TP, Carrara P, Altamura E, D'Aguanno E, Caputo M, Luisi PL, Mavelli F: **Recent biophysical issues about the preparation of solute-filled lipid vesicles**. *Mech. Adv. Mater. Struct.* 2014, **22**:748–759.
196. Israelachvili JN: **Electrostatic forces between surfaces in liquids**. In *Intermolecular and Surface Forces*. Edited by Israelachvili JN. San Diego: Academic Press, 2011: 291–340.
197. Kegel WK, van der Schoot P: **Competing hydrophobic and screened-coulomb interactions in hepatitis B virus capsid assembly.** *Biophys. J.*

2004, **86**:3905–3913.
198. Griesbeck O, Baird GS, Campbell RE, Zacharias DA, Tsien RY: **Reducing the environmental sensitivity of yellow fluorescent protein. Mechanism and applications.** *J. Biol. Chem.* 2001, **276**:29188–29194.
199. Min J, Kim S, Lee J, Kang S: **Lumazine synthase protein cage nanoparticles as modular delivery platforms for targeted drug delivery**. *RSC Adv.* 2014, **4**:48596–48600.
200. Malyutin AG, Dragnea B: **Budding pathway in the templated assembly of viruslike particles**. *J. Phys. Chem. B* 2013, **117**:10730–10736.
201. Uetrecht C, Watts NR, Stahl SJ, Wingfield PT, Steven AC, Heck AJR: **Subunit exchange rates in Hepatitis B virus capsids are geometry- and temperature-dependent.** *Phys. Chem. Chem. Phys.* 2010, **12**:13368–13371.
202. Teschke CM, McGough A, Thuman-Commike PA: **Penton release from P22 heat-expanded capsids suggests importance of stabilizing penton-hexon interactions during capsid maturation.** *Biophys. J.* 2003, **84**:2585–2592.
203. MacBeath G, Kast P, Hilvert D: **Probing enzyme quaternary structure by combinatorial mutagenesis and selection.** *Protein Sci.* 1998, **7**:1757–67.
204. Bhattar SL, Kolekar GB, Patil SR: **Fluorescence resonance energy transfer between perylene and riboflavin in micellar solution and analytical application on determination of vitamin B2**. *J. Lumin.* 2008, **128**:306–310.
205. Zhang X, Konarev P V, Petoukhov M V, Svergun DI, Xing L, Cheng RH, Haase I, Fischer M, Bacher A, Ladenstein R, et al.: **Multiple assembly states of lumazine synthase: a model relating catalytic function and molecular assembly.** *J. Mol. Biol.* 2006, **362**:753–70.
206. Bacher A, Ludwig HC, Schnepple H, Ben-Shaul Y: **Heavy riboflavin synthase from Bacillus subtilis. Quarternary structure and reaggregation**. *J. Mol. Biol.* 1986, **187**:75–86.
207. Sharon M: **Structural MS pulls its weight**. *Science* 2013, **340**:1059–1060.
208. Heck AJR: **Native mass spectrometry: a bridge between interactomics and structural biology.** *Nat. Methods* 2008, **5**:927–933.
209. Fuerstenau SD, Benner WH, Thomas JJ, Brugidou C, Bothner B, Siuzdak G: **Mass spectrometry of an intact virus**. *Angew. Chemie* 2001, **113**:1011–1011.
210. Tito MA, Tars K, Valegård K, Hajdu J, Robinson C V: **Electrospray time of flight mass spectrometry of the intact MS2 virus capsid**. *J. Am. Chem. Soc.* 2000, **122**:3550–3551.
211. Uetrecht C, Barbu IM, Shoemaker GK, van Duijn E, Heck AJR: **Interrogating viral capsid assembly with ion mobility–mass spectrometry**. *Nat. Chem.* 2011, **3**:126–132.

212. Snijder J, Uetrecht C, Rose RJ, Sanchez-Eugenia R, Marti GA, Agirre J, Guérin DMA, Wuite GJL, Heck AJR, Roos WH: **Probing the biophysical interplay between a viral genome and its capsid.** *Nat. Chem.* 2013, **5**:502–9.
213. Brasch M, de la Escosura A, Ma Y, Uetrecht C, Heck AJR, Torres T, Cornelissen JJLM: **Encapsulation of phthalocyanine supramolecular stacks into virus-like particles.** *J. Am. Chem. Soc.* 2011, **133**:6878–6881.
214. Pierson EE, Keifer DZ, Selzer L, Lee LS, Contino NC, Wang JCY, Zlotnick A, Jarrold MF: **Detection of late intermediates in virus capsid assembly by charge detection mass spectrometry**. *J. Am. Chem. Soc.* 2014, **136**:3536–3541.
215. Keifer DZ, Pierson EE, Hogan JA, Bedwell GJ, Prevelige PE, Jarrold MF: **Charge detection mass spectrometry of bacteriophage P22 procapsid distributions above 20 MDa**. *Rapid Commun. Mass Spectrom.* 2014, **28**:483–488.
216. Rurup WF, Snijder J, Koay MST, Heck AJR, Cornelissen JJLM: **Self-sorting of foreign proteins in a bacterial nanocompartment.** *JACS* 2014, **136**:3828–3832.
217. Snijder J, Rose RJ, Veesler D, Johnson JE, Heck AJR: **Studying 18 MDa virus assemblies with native mass spectrometry**. *Angew. Chemie - Int. Ed.* 2013, **52**:4020–4023.
218. Lemaire D, Marie G, Serani L, Laprévote O: **Stabilization of gas-phase noncovalent macromolecular complexes in electrospray mass spectrometry using aqueous triethylammonium bicarbonate buffer**. *Anal. Chem.* 2001, **73**:1699–1706.
219. Harms ZD, Selzer L, Zlotnick A, Jacobson SC: **Monitoring assembly of virus capsids with nanofluidic devices**. *ACS Nano* 2015, **9**:9087–9096.
220. Jäckel C, Hilvert D: **Biocatalysts by evolution**. *Curr. Opin. Biotechnol.* 2010, **21**:753–759.
221. Urlacher VB, Girhard M: **Cytochrome P450 monooxygenases: an update on perspectives for synthetic application**. *Trends Biotechnol.* 2012, **30**:26–36.
222. Romero PA, Arnold FH: **Exploring protein fitness landscapes by directed evolution**. *Nat. Rev. Mol. Cell Biol.* 2009, **10**:866–876.
223. Nagai K, Thøgersen HC: **Generation of β-globin by sequence-specific proteolysis of a hybrid protein produced in Escherichia coli**. *Nature* 1984, **309**:810–812.
224. Yildiz I, Shukla S, Steinmetz NF: **Applications of viral nanoparticles in medicine**. *Curr. Opin. Biotechnol.* 2011, **22**:901–908.
225. Manchester M, Steinmetz NF: **Viral Nanoparticles: Tools for Materials**

Science & Biomedicine. Singapore: Pan Stanford Publishing, 2011.
226. The UniProt Consortium: **UniProtKB - O66529 (RISB_AQUAE)**. Website 2015, http://www.uniprot.org/uniprot/O66529.
227. Azuma Y, Zschoche R, Tinzl M, Hilvert D: **Quantitative Packaging of Active Enzymes into a Protein Cage**. *Angew. Chemie Int. Ed.* 2016, **55**:1531–1534.
228. Sasso S, Ramakrishnan C, Gamper M, Hilvert D, Kast P: **Characterization of the secreted chorismate mutase from the pathogen Mycobacterium tuberculosis.** *FEBS J.* 2005, **272**:375–389.
229. Brouwer AM: **Standards for photoluminescence quantum yield measurements in solution (IUPAC technical report)**. *Pure Appl. Chem.* 2011, **83**:2213–2228.
230. Conway JH, Sloane NJA: **Sphere packings, lattices and groups**. New York: Springer, 1999.
231. McCann JJ, Choi UB, Zheng L, Weninger K, Bowen ME: **Optimizing methods to recover absolute FRET efficiency from immobilized single molecules**. *Biophys. J.* 2010, **99**:961–970.
232. Fersht A: **Structure and mechanism in protein science: A guide to enzyme catalysis and protein folding**. New York: W. H. Freeman and Co., 1999.
233. DiMaio F, Leaver-Fay A, Bradley P, Baker D, André I: **Modeling symmetric macromolecular structures in Rosetta3.** *PLoS One* 2011, **6**:e20450.
234. Leaver-Fay A, Tyka M, Lewis SM, Lange OF, Thompson J, Jacak R, Kaufman KW, Renfrew PD, Smith CA, Sheffler W, et al.: **Rosetta3: An object-oriented software suite for the simulation and design of macromolecules**. *Methods Enzymol.* 2011, **487**:545–574.
235. Comellas-Aragonès M, Engelkamp H, Claessen VI, Sommerdijk NAJM, Rowan AE, Christianen PCM, Maan JC, Verduin BJM, Cornelissen JJLM, Nolte RJM: **A virus-based single-enzyme nanoreactor**. *Nat. Nanotechnol.* 2007, **2**:635–9.
236. Minten IJ, Hendriks LJA, Nolte RJM, Cornelissen JJLM: **Controlled encapsulation of multiple proteins in virus capsids.** *JACS* 2009, **131**:17771–17773.
237. Fiedler JD, Brown SD, Lau JL, Finn MG: **RNA-directed packaging of enzymes within virus-like particles**. *Angew. Chem., Int. Ed.* 2010, **49**:9648–9651.
238. O'Neil A, Reichhardt C, Johnson B, Prevelige PE, Douglas T: **Genetically programmed in vivo packaging of protein cargo and its controlled release from bacteriophage P22**. *Angew. Chem., Int. Ed.* 2011, **50**:7425–8.
239. Glasgow JE, Capehart SL, Francis MB, Tullman-Ercek D: **Osmolyte-mediated encapsulation of proteins inside MS2 viral capsids**. *ACS Nano* 2012, **6**:8658–8664.

240. Glasgow JE, Asensio MA, Jakobson CM, Francis MB, Tullman-Ercek D: **Influence of electrostatics on small molecule flux through a protein nanoreactor**. *ACS Synth. Biol.* 2015, **4**:1011–1019.
241. Schmidt J, Ehasz C, Epperson M, Klas K, Wyatt J, Hennig M, Forconi M: **The effect of the hydrophobic environment on the retro-aldol reaction: comparison to a computationally-designed enzyme.** *Org. Biomol. Chem.* 2013, **11**:8419–25.
242. Woods AS, Ferré S: **Amazing stability of the arginine-phosphate electrostatic interaction.** *J. Proteome Res.* 2005, **4**:1397–402.
243. Park DH, Plapp B V: **Isoenzymes of horse liver alcohol dehydrogenase active on ethanol and steroids. cDNA cloning, expression, and comparison of active sites.** *J. Biol. Chem.* 1991, **266**:13296–302.
244. Strasser F, Huyng MN, Plapp B V.: **Activity of liver alcohol dehydrogenases on steroids**. In *Enzymology and Molecular Biology of Carbonyl Metabolism 6*. Edited by Weiner H, Lindahl R, Crabb DW, Flynn TG. New York: Plenum Press, 1996: 313–320.
245. Giacomini D, Galletti P, Quintavalla A, Gucciardo G, Paradisi F: **Highly efficient asymmetric reduction of arylpropionic aldehydes by Horse Liver Alcohol Dehydrogenase through dynamic kinetic resolution**. *Chem. Commun.* 2007, **Epub**:4038.
246. Galletti P, Emer E, Gucciardo G, Quintavalla A, Pori M, Giacomini D: **Chemoenzymatic synthesis of (2S)-2-arylpropanols through a dynamic kinetic resolution of 2-arylpropanals with alcohol dehydrogenases.** *Org. Biomol. Chem.* 2010, **8**:4117–23.
247. Quaglia D, Irwin JA, Paradisi F: **Horse liver alcohol dehydrogenase: new perspectives for an old enzyme.** *Mol. Biotechnol.* 2012, **52**:244–50.
248. Nagel ZD, Cun S, Klinman JP: **Identification of a long-range protein network that modulates active site dynamics in extremophilic alcohol dehydrogenases.** *J. Biol. Chem.* 2013, **288**:14087–97.
249. Frey R, Mantri S, Rocca M, Hilvert D: **Bottom-up Construction of a Primordial Carboxysome Mimic**. *J. Am. Chem. Soc.* 2016, doi:10.1021/jacs.6b04744.
250. Obexer R, Godina A, Garrabou X, Mittl PRE, Baker D, Griffiths AD, Hilvert D: **Emergence of a catalytic tetrad during evolution of a highly efficient artificial aldolase**. *submitted* [date unknown], [no volume].
251. Shrivastava R, Basu A, Phale PS: **Purification and characterization of benzyl alcohol- and benzaldehyde- dehydrogenase from Pseudomonas putida CSV86**. *Arch. Microbiol.* 2011, **193**:553–563.
252. Paliwal V, Raju SC, Modak A, Phale PS, Purohit HJ: **Pseudomonas putida**

CSV86: A candidate genome for genetic bioaugmentation. *PLoS One* 2014, **9**:1–12.

253. Jo JE, Mohan RS, Rathnasingh C, Selvakumar E, Jung WC, Park S: **Cloning, expression, and characterization of an aldehyde dehydrogenase from Escherichia coli K-12 that utilizes 3-hydroxypropionaldehyde as a substrate**. *Appl. Microbiol. Biotechnol.* 2008, **81**:51–60.
254. Stachow CS, Stevenson IL, Day D: **Purification and properties of nicotinamide adenine dinucleotide phosphate-specific benzaldehyde dehydrogenase from Pseudomonas**. *J. Biol. Chem.* 1967, **242**:5294–5300.
255. MacKintosh RW, Fewson CA: **Benzyl alcohol dehydrogenase and benzaldehyde dehydrogenase II from Acinetobacter calcoaceticus. Substrate specificities and inhibition studies.** *Biochem. J.* 1988, **255**:653–61.
256. Fan C, Cheng S, Liu Y, Escobar CM, Crowley CS, Jefferson RE, Yeates TO, Bobik TA: **Short N-terminal sequences package proteins into bacterial microcompartments**. *Proc. Natl. Acad. Sci.* 2010, **107**:7509–7514.
257. Fan C, Cheng S, Sinha S, Bobik TA: **Interactions between the termini of lumen enzymes and shell proteins mediate enzyme encapsulation into bacterial microcompartments**. *Proc. Natl. Acad. Sci.* 2012, **109**:14995–15000.
258. Fan C, Bobik TA: **The N-terminal region of the medium subunit (PduD) packages adenosylcobalamin-dependent diol dehydratase (PduCDE) into the Pdu microcompartment**. *J. Bacteriol.* 2011, **193**:5623–5628.
259. Jakobson CM, Kim EY, Slininger MF, Chien A, Tullman-Ercek D: **Localization of proteins to the 1,2-propanediol utilization microcompartment by non-native signal sequences is mediated by a common hydrophobic motif**. *J. Biol. Chem.* 2015, **290**:24519–24533.
260. Liao DI, Wawrzak Z, Calabrese JC, Viitanen P V., Jordan DB: **Crystal structure of riboflavin synthase**. *Structure* 2001, **9**:399–408.
261. Gerhardt S, Schott A-K, Kairies N, Cushman M, Illarionov B, Eisenreich W, Bacher A, Huber R, Steinbacher S, Fischer M: **Studies on the reaction mechanism of riboflavin synthase: X-ray crystal structure of a complex with 6-carboxyethyl-7-oxo-8-ribityllumazine**. *Structure* 2002, **10**:1371–1381.
262. Serer MI, Bonomi HR, Guimaraes BG, Rossi RC, Goldbaum FA, Klinke S: **Crystallographic and kinetic study of riboflavin synthase from Brucella abortus, a chemotherapeutic target with an enhanced intrinsic flexibility**. *Acta Crystallogr.D.Biol Crystallogr.* 2014, **70**:1419–1434.
263. Dames SA, Kammerer RA, Wiltscheck R, Engel J, Alexandrescu AT: **NMR**

structure of a parallel homotrimeric coiled coil. *Nat. Struct. Biol.* 1998, **5**:687–91.

264. Pédelacq J-D, Cabantous S, Tran T, Terwilliger TC, Waldo GS: **Engineering and characterization of a superfolder green fluorescent protein.** *Nat. Biotechnol.* 2006, **24**:79–88.
265. Bacher A, Baur R, Eggers U, Harders HD, Otto MK, Schnepple H: **Riboflavin synthases of Bacillus subtilis. Purification and properties**. *J. Biol. Chem.* 1980, **255**:632–637.
266. Woolfson DN: **The design of coiled-coil structures and assemblies**. *Adv. Protein Chem.* 2005, **70**:79–112.
267. Bromley EHC, Channon K, Moutevelis E, Woolfson DN: **Peptide and protein building blocks for synthetic biology: from programming biomolecules to self-organized biomolecular systems**. *ACS Chem. Biol.* 2008, **3**:38–50.
268. Boyle AL, Woolfson DN: **De novo designed peptides for biological applications**. *Chem. Soc. Rev.* 2011, **40**:4295–306.
269. Woolfson DN, Bartlett GJ, Bruning M, Thomson AR: **New currency for old rope: from coiled-coil assemblies to α-helical barrels.** *Curr. Opin. Struct. Biol.* 2012, **22**:432–41.
270. Nautiyal S, Woolfson DN, King DS, Alber T: **A designed heterotrimeric coiled coil**. *Biochemistry* 1995, **34**:11645–51.
271. Nautiyal S, Alber T: **Crystal structure of a designed, thermostable, heterotrimeric coiled coil.** *Protein Sci.* 1999, **8**:84–90.
272. Schulenburg C, Hilvert D: **Protein conformational disorder and enzyme catalysis**. *Top. Curr. Chem.* 2013, **337**:41–67.
273. Brauchle M, Hansen S, Caussinus E, Lenard A, Ochoa-Espinosa A, Scholz O, Sprecher SG, Plückthun A, Affolter M: **Protein interference applications in cellular and developmental biology using DARPins that recognize GFP and mCherry.** *Biol. Open* 2014, **3**:1252–61.
274. Putri RM, Cornelissen JJLM, Koay MST: **Self-assembled cage-like protein structures**. *ChemPhysChem* 2015, **16**:911–8.
275. Canny G, Levy O: **Bactericidal/permeability-increasing protein (BPI) and BPI homologs at mucosal sites**. *Trends Immunol.* 2008, **29**:541–7.
276. Kostiainen MA, Hiekkataipale P, Laiho A, Lemieux V, Seitsonen J, Ruokolainen J, Ceci P: **Electrostatic assembly of binary nanoparticle superlattices using protein cages**. *Nat. Nanotechnol.* 2013, **8**:52–56.
277. Liljeström V, Mikkilä J, Kostiainen MA: **Self-assembly and modular functionalization of three-dimensional crystals from oppositely charged proteins.** *Nat. Commun.* 2014, **5**:4445–53.