Functionalizing Capsid Forming Proteins

A dissertation submitted to
ETH Zürich

for the degree of
Doctor of Sciences

presented by

Bigna Wörsdörfer
Dipl.-Biochem. University of Bayreuth
born on May 22, 1980
citizen of Germany

Accepted on the recommendation of
Prof. Dr. Donald Hilvert, examiner
Prof. Dr. Nenad Ban, co-examiner

Zürich 2010
Acknowledgement

I am grateful to all the people who supported me during my PhD thesis with scientific or personal advice, friendship and love.

I want to thank my PhD supervisor Prof. Donald Hilvert for always giving me helpful advice to further the project whenever I needed it and yet allowing me the freedom to develop my project and grow as an independent researcher. I have profited greatly from his vast knowledge and his wonderful way of writing.

I thank Prof. Nenad Ban for being co-examiner of my thesis.

I am grateful to Prof. Kenneth Woycechowsky, who together with Prof. Florian Seebeck conceived the exciting lumazine synthase projects. Ken introduced me, experimentally and theoretically, to the project and taught me a lot about writing, and making posters and presentations. His enthusiasm for science and sharing the emotions with him after having received good or bad results motivated me. I also want to thank him for taking the unforgettable night shifts to measure the growth curves.

Many thanks go to Dr. Peter Kast for all of his advice and his help especially with molecular biological problems.

Dr. Martin Neuenschwander taught me many molecular biological methods and I am thankful having learned a lot from him regarding designing and carrying out experiments. Further, I thank Dr. Andreas Kleeb for teaching me much about performing directed evolution, and Dr. Joris Beld for his help with various things including Linux, the figures in the introduction of this thesis and installation of the programs I needed to evaluate the cryo-EM data.
I also thank my students Davide Speziga, Sylvain Rütti and Lisa Henning for their participation and interest in my research. With her work, Lisa made an important contribution to the aldolase project described in Chapter 5.

I want to thank Dr. Takashi Ishikawa and Dr. Miriam Lucas at the Electron Microscopy Center of the ETH Zürich (EMEZ) for collecting the EM images and assistance with EM data analysis, and Dr. Bernd Roschitzki at the Functional Genomics Center Zürich for collection of mass spectrometric data and assistance with data analysis.

I enjoyed many fruitful discussions with Prof. Hans-Martin Fischer about stress response and I want to thank him for giving me the opportunity to carry out the RNA work in his and Prof. Hauke Hennecke’s lab. I am very thankful to Simona Huwiler, who showed me how to isolate and work with RNA with great commitment. I also want to thank Prof. Franz Narberhaus for contributive discussions on the stress response and the heat shock protein ibpA.

I want to thank all the members of the Hilvert group for their advice, help with experiments, valuable discussions, and for the nice time in and outside of the lab. Many of them became friends and we have had a great time and many wonderful experiences together.

Special thanks go to all my past and present F336 lab members for the always pleasant atmosphere, lots of personal and scientific advice, and friendship. Sometimes the lab felt like an isle of retreat.

I am grateful to my family and Jörg for their unconditional support, encouragement and love, which made me stronger in hard times and happier in happy times. They have always been there for me.
Parts of this thesis have been published

Directed evolution of a protein container
B. Wörsdörfer, K. J. Woycechowsky, D. Hilvert
Science 331, 589 (2011)
# Table of Contents

Abstract

Zusammenfassung 5

1 Introduction 9

1.1 Symmetric oligomers 9
1.2 Protein capsids are highly symmetric oligomers 13
1.3 Functions of protein capsids in nature 15
1.4 Engineering protein capsids for bionanotechnology 18
1.5 Lumazine synthase 26
1.6 Aims of this thesis 28

2 Directed evolution of a protein container 31

2.1 Introduction 31
2.2 Results 32

2.2.1 Encapsulation of HIV protease by AaLS-neg 32
2.2.2 Directed evolution 36
2.2.3 Characterization of the evolved variant AaLS-13 40
2.3 Discussion 49

3 Properties of an evolved protein container 51

3.1 Introduction 51
3.2 Results 52

3.2.1 In vivo guest encapsulation 52
3.2.2 Stability of the capsids and host-guest complexes 59
3.3 Discussion 66

4 In vitro encapsulation of proteins 69

4.1 Introduction 69
4.2 Results 71
4.3 Discussion 77
### 5 Harnessing symmetry for the de novo creation of enzyme activity

#### 5.1 Introduction

#### 5.2 Results

- 5.2.1 Directed evolution with in vivo selection systems
- 5.2.2 Rational design of aldolase catalysts

#### 5.3 Discussion

### 6 Perspective

### 7 Materials and Methods

#### 7.1 Materials and standard methods

#### 7.2 Plasmid construction

#### 7.3 Library construction

#### 7.4 Selections

#### 7.5 Growth tests

#### 7.6 Monitoring the stress response by real-time PCR

#### 7.7 Protein production and purification

#### 7.8 Detection of HIV protease

#### 7.9 Mixing experiments

#### 7.10 Electron microscopy (EM)

#### 7.11 Mass spectrometry

#### 7.12 Circular dichroism (CD) spectroscopy

#### 7.13 Encapsulation efficiencies for GFP variants

#### 7.14 Dissociation of AaLS-guest complexes

#### 7.15 In vitro loading

#### 7.16 Kinetic measurements for retro-aldol cleavage of (+)-methodol

#### 7.17 96 well plate screening assay for retro-aldol cleavage of (+)-methodol

### 8 References

### Curriculum Vitae
Compartmentalization in nature organizes cellular processes and metabolic pathways and enables Darwinian evolution by linking genotype and phenotype. Whereas lipid membrane-bound organelles are generally used in eukaryotes to achieve sequestration, for many specialized functions nature employs protein containers. Protein capsids are used to compartmentalize enzymes and whole metabolic processes, to store iron, to mediate the transport and transfer of viral genomes, and to assist protein folding.

There is considerable interest in mimicking nature and exploiting the advantages of encapsulation for various medical, and bio- and nanotechnological applications. Engineered protein capsids have been used as templates for the controlled synthesis of nanomaterials, as gene and drug delivery vehicles, and towards bio- and medical imaging. The confinement of catalysts inside protein shells has been used to control catalysis, and it often has a stabilizing effect on the guest, and allows enzymes to be studied in defined environments.

In this thesis, the natural capsid-forming protein lumazine synthase (LS) has been reengineered for new functions. The natural function of LS is to catalyze the penultimate step in the biosynthesis of riboflavin, but it has been shown that LS can also serve as scaffold for protein engineering purposes outside the cell. An important prerequisite for the redesign of proteins is the ability of their structure to tolerate multiple changes. The pentameric *Saccharomyces cerevisiae* LS (ScLS) has been shown to retain quaternary structure upon extensive mutation of tunnel residues. The icosahedral capsids of LS present an attractive alternative to other natural protein containers that can be exploited for bio- and nanotechnological purposes. This was demonstrated in a study that used LS from *Bacillus subtilis* for the size-constrained synthesis of monodisperse iron oxide nanoparticles inside the LS cavity, and by the engineering of an protein encapsulation system based on LS capsids from *Aquifex aeolicus* (AaLS). Encapsulation of GFP was achieved by engineered electrostatic interactions between the negatively charged luminal surface of the lumazine synthase variant AaLS-neg and a positively charged deca-
arginine (R$_{10}$) tag appended to the guest. Here, the optimization and extension of this rationally designed AaLS encapsulation system is described.

In Chapter 2 the ability of AaLS capsids to serve as novel nanocompartments was optimized by directed evolution. We transferred the R$_{10}$-tag to HIV protease – a toxic enzyme when produced cytoplasmically – and thereby sequestered the protease within AaLS-neg capsids. The protein container acts as a primitive prokaryotic organelle, protecting the *E. coli* host from the lethal effects of the protease by physically blocking its access to potential substrates. The resulting growth advantage enabled directed evolution of improved capsids. After four rounds of mutagenesis and selection, we obtained a AaLS variant, AaLS-13, with a five- to ten-fold higher loading capacity than the starting capsid, which permitted efficient growth even at high intracellular concentrations of HIV protease. The superior properties of the evolved capsid can be ascribed to multiple mutations that increase the net negative charge on its luminal surface and thereby enhance engineered Coulombic interactions between host and guest.

Chapter 3 describes the characterization of this evolved encapsulation system and the encapsulation of alternative guests. Determining the specificity of guest loading in vivo revealed that AaLS-13 copurifies with very abundant and highly positively charged ribosomal proteins. Nevertheless, the preferential incorporation of tagged HIV protease suggests the system exhibits selectivity towards the intended guest. Encapsulation of GFP variants possessing different net charge and charge distributions demonstrates that AaLS-13 can be efficiently loaded in vivo with other proteins possessing sufficient net positive charge. Structural investigations showed that AaLS-13 capsids can disassemble into pentamers in the absence of a guest, and that guest loading with HIV protease-R$_{10}$ stabilizes the AaLS-13 capsid. The host-guest complexes formed between AaLS-13 and highly positively charged proteins are then very stable, since no dissociation of guest was seen even after many days.

In Chapter 4 we show that the structural transitions and dynamics of AaLS-13 can be utilized for controlled disassembly and reassembly to efficiently encapsulate a positively charged GFP variant in vitro. Intact capsids with extremely high guest loadings of up to 60 GFP molecules per capsid could be obtained under favorable conditions. This loading level is about four times higher than that obtained upon
coproduction of host and guest in vivo. These results suggest that controlled and specific in vitro encapsulation of a variety of protein and non-protein guests into AaLS-13 should be possible.

The interior of AaLS capsids is connected to the bulk surrounding by tunnels located at the fivefold symmetry axis of the pentameric building blocks. In Chapter 5 efforts to endow the tunnel of the pentameric LS variant from *S. cerevisiae* with de novo catalytic activity are described. Directed evolution using powerful in vivo selection systems for chorismate mutase, prephenate dehydratase, and dihydrofolate reductase activities did not yield any viable catalysts. However, rational design of tunnel variants for the well-understood and less demanding retro-aldol reaction according to simple mechanistic design principles imparted a tunnel variant with aldolase activity. In the designs, multiple lysine residues were introduced into the tunnel to afford a nucleophilic amine for catalysis. One design was particularly active and exhibited rate accelerations of 2500 over the uncatalyzed reaction. Its activity can be ascribed to imine formation with one of the new lysines, which has a depressed pK_a due to its proximity to the other symmetry related (and protonated) lysine residues. Applying directed evolution and computational design should allow improving the catalytic activity of the LS tunnel variant.

The apparent ease with which LS capsids and its pentameric building blocks could be functionalized underscores the considerable potential of these structures for diverse nano- and biotechnological applications. Building on the results obtained in this thesis the engineering of AaLS based nanocompartments with novel functions should be straightforward. Exciting future applications to pursue include the development of drug and gene delivery systems, host-guest complexes for bio- and medical imaging, and the construction of multienzyme complexes that perform whole reaction sequences by compartmentalizing several catalysts. Functionalized pores could serve as additional catalysts and as selective gates to control passage of molecules across the protein shell.


Komplex zwischen AaLS-13 and hoch positiv geladenen Proteinen ist dann sehr stabil und selbst nach mehreren Tagen kann keine Dissoziation des Gasts beobachtet werden.


Die ersichtliche Leichtigkeit, mit der LS Kapside und dessen pentamerische Bausteine funktionalisiert werden konnten, unterstreicht das erhebliche Potential dieser Strukturen für diverse nano- und biotechnologische Anwendungen.
1 INTRODUCTION

1.1 Symmetric oligomers

The majority of soluble and membrane proteins, and nearly all structural proteins form symmetrical oligomeric, particularly homooligomeric, complexes. It is estimated that 50% to 70% of proteins with known structure are homooligomers, and almost all of them are symmetric (1-3). Hence, there must be some selective advantage driving the evolution toward such structures.

The evolution of symmetric, oligomeric complexes is commonly thought to have arisen from structural and functional driving forces (1, 4). Koshland speculated that cells early in evolution needed large proteins to reduce their loss through leaky membranes (5). Because mechanisms had not yet been developed for active transport, these cells contained large pores to allow diffusion of molecules into and out of the cell. In modern cells sophisticated transport mechanisms eliminate this rationale. However, large proteins still have many advantages today. The morphological functions of structural proteins like the long actin and tubulin filaments require large structures. The same is true for ring-forming systems and hollow capsids such as viral shells. But in general, large proteins are believed to be more stable due to extensive internal interactions, better enthalpic compensation of entropy, and reduced exposed surface area. They also provide enough room for the incorporation of regulatory elements.

Larger proteins are advantageous compared to smaller proteins. Nevertheless, they have a greater biosynthetic cost. This trade-off is optimized by nature through the construction of large assemblies from many identical subunits. Homooligomers make efficient use of genetic information since only genetic coding of a small subunit is required. This coding economy is especially important for organisms with limited genome size like viruses (6). For higher organisms error control might be a more essential reason for using oligomers. The chances of errors occurring in the synthesis of proteins increase with the size of the protein. Assuming an average frequency of missense error – namely, that incorporation of a wrong amino acid at a given position in the sequence occurs with a frequency of ca. $5 \times 10^{-4}$ per codon,
about one fourth of the population of a 500 amino acid protein will have a mutation. Proteins with 2000 amino acids nearly always have an error (7). Missense errors are usually tolerated fairly well, but processivity errors, in which translation is terminated prematurely, have a more severe impact for the organism. The frequency of premature translation termination is estimated to be $2.5 \times 10^{-4}$, leading to loss of 12.5% for a 500 amino acid protein. A protein with 3000 amino acids will only rarely be translated completely. In addition, the assembly of large complexes from small subunits allows small subunits with defects to be discarded at much lower biosynthetic cost than large proteins.

Oligomerization additionally provides attractive opportunities for regulation. Cooperative interactions between the subunits in an oligomer enable allosteric regulation, which is especially useful when information needs to be passed efficiently from one subunit to all the subunits of the protein. The tetrameric protein hemoglobin is a prime example of such allosteric control. Binding of oxygen to one subunit induces a conformational change that enhances the oxygen affinity in all the remaining subunits. Another regulatory feature of oligomers is the possibility of using the assembly process itself to control the activity of the oligomer or its structure.

Oligomers can have multiple binding sites, which lets them serve as protein crosslinkers. Actin-bundling proteins that link actin filaments are good examples. Multivalent binding of an oligomer can also increase the binding strength to a target receptor that contains multiple binding sites, since entropy loss upon binding is reduced compared to binding of numerous isolated monomers. Once one site has bound to the target, the other sites are held in close proximity, facilitating their binding to the other sites.

Almost all protein oligomers found in nature are highly symmetric. Goodsell and Olson noted that symmetry is the rule rather than the exception for proteins (1). Several explanations for the predominance of symmetry in protein complexes have been formulated. First, analyses of simple models and computational modeling have shown that symmetric protein-protein interfaces tend to be energetically favorable. Cornish-Bowden and Koshland analyzed the thermodynamic basis of arranging four subunits in a plane (8). Assuming two binding surfaces, defined as P and Q, four different arrangements are possible (Fig. 1.1). Two of these are
symmetric with either P-P and Q-Q interfaces, or P-Q interfaces, while the other two complexes are asymmetric containing P-P, P-Q and Q-Q interactions. In this model even small differences in the binding energies between the P-P and Q-Q and the P-Q interactions are sufficient to strongly favor the symmetric assemblies over the asymmetric ones. Since the binding strength of the different types of interfaces will likely differ considerably in natural proteins, the asymmetric species, which have all three types of binding interactions, will rarely occur.

Figure 1.1. Possible planar arrangements for a tetramer using two binding surfaces, defined as P and Q. Gray subunits are flipped compared to white subunits. The upper two complexes are symmetric whereas the lower two are asymmetric.

Baker et al. have used computational docking calculations on protein models to explain the high abundance of symmetrical homooligomers found in nature (9). They randomly docked homodimers and found that the fraction of symmetric complexes rises with increasingly stringent selection towards low-energy structures. In the population of all complexes that have a sufficient binding energy to overcome the entropic costs of dimerization only 0.02% are symmetric. The fraction of symmetric assemblies increases to more than 50% in the very low-energy subpopulations. These results suggest that symmetric assemblies can arise solely through selection of complexes with a sufficiently high interaction energy. As a consequence, symmetric complexes are overpopulated in initial pools of
randomly arising assemblies and therefore represent more accessible targets for natural evolution.

Homooligomers can assemble by using identical binding surfaces to build the interface. Such a process has been called isologous association by Monod et al. and corresponds to formation of the P-P or Q-Q interfaces in the example above. A dimer with such an isologous interface necessarily forms a closed symmetric structure where further assembly is only possible by using a new binding site. Again employing isologous association for this new binding site, a closed tetramer is formed, which has dihedral symmetry (D2 symmetry). When different binding surfaces are used to form the interface, namely the P-Q interfaces or heterologous association, open and closed structures can be formed in principle. However, closed structures giving rise to cyclic symmetries contain the maximum number of stabilizing intersubunit interactions and they should therefore be more stable. For example, four subunits in an open extended chain will have only three stabilizing interfaces with two less stable subunits at the ends, whereas a closed cyclic structure will have four favorable interactions.

Symmetrical closed packing not only provides higher stability, it also results in finite assemblies with a defined number of subunits. This is important since function might depend on the assembly state. Unwanted protein aggregation, which can result in severe diseases such as Alzheimer’s and prion diseases, must be avoided. Helical symmetries are also possible, but because they are not bounded, special terminating mechanisms are required.

Other benefits of symmetry that have been proposed include a higher folding efficiency due to a more favorable energy landscape (10). Symmetry played a central role in the development of the concept of allostERIC regulation by Monod, Wyman and Changeux (4). They also argued that symmetric assemblies more easily emerge and evolve. When two monomers associate as a dimer, the favorable interaction of any pair of complementary residues, e.g. a salt bridge, results in the symmetry-related formation of a second stabilizing interaction. In this way mutation of one residue in such a primitive dimer that improves the association yields two beneficial interfacial interactions. The effect of the mutation is thus multiplied due to symmetry. In analogy, the structural and functional effects of one mutation in subunits of a \( C_n \) symmetric oligomer, resulting in \( n \) symmetry-related
mutations, should be greatly amplified. Thus, the region close to the symmetry axis should be particularly sensitive to changes. In the words of Monod:

“Because of the inherent cooperativity of their structure, symmetrical oligomers should constitute particularly sensitive targets for molecular evolution, allowing much stronger selective pressures to operate in the random pursuit of functionally adequate structures” (4).

As this analysis attests, large oligomers have many advantages over small monomers. The predominance of symmetry among existing protein complexes can be explained by the enrichment of symmetric primitive complexes in the initial pool, due to more stable assembly and greater ease of emergence, and by a combination of other beneficial factors including those described above. However, functional needs are still the strongest evolutionary driving force. To accomplish specific functions protein complexes often break from perfect symmetry, resulting in quasisymmetry (identical subunits that adopt similar but different positions), pleomorphism (identical subunits that assemble to different complexes), pseudosymmetry (different subunits that form approximately symmetrical structures), and symmetry mismatches (interaction of components with different symmetry). Asymmetry at the atomic level is common and nearly all complexes show local asymmetry in side chain conformations. However, global asymmetry is rarely observed.

1.2 Protein capsids are highly symmetric oligomers

Nature uses cubic symmetry to assemble large protein containers from small, often identical subunits for specialized biological functions that require compartmentalization. Three types of cubic symmetries exist: tetrahedral, octahedral, and icosahedral. Ferritin, an iron storage capsid, is an example of octahedral symmetry, whereas most virus capsids and other protein containers possess icosahedral symmetry. To assemble a capsid of given size using icosahedral symmetry is most economic since it requires the least number of subunits (11). Perfect icosahedra are built of 60 identical subunits arranged as twelve pentamers.
More than 60 identical subunits cannot reside in exactly equivalent positions in an icosahedron. To build larger capsids nature uses quasisymmetry where the identical subunits are placed in slightly different environments. To describe the structures of capsids with more than 60 subunits and icosahedral symmetry Caspar and Klug developed the concept of quasi-equivalence and the triangulation number \( T \), and showed that icosahedral capsids can be build from pentamers and hexamers. One subunit is either part of a pentamer or a hexamer and adjoins to other pentamers or hexamers. As a result, the subunits are in quasi-equivalent positions. By constructing capsids with icosahedral symmetry from a lattice of regular triangles they introduced the triangulation number \( T \), where \( T = h^2 + hk + k^2 \), and \( h \) and \( k \) are integers. \( T \) indicates the number of nonequivalent subunit positions in the capsids. Icosahedral capsids are assembled from 60\( T \) subunits that form 12 pentamers and 10\( (T-1) \) hexamers.

To gain insight into the question as to why protein capsids most often adopt icosahedral symmetry, calculations on simplified capsid models have been performed. In an approach using Monte Carlo simulations, Zandi et al. found that icosahedral structures, built as described by Caspar and Klug, are highly abundant because their structures are minimal in free energy (12). The capsid models used in this study were constructed from pentameric and hexameric building blocks, called capsomers. In the Monte Carlo simulations these capsomers were allowed to interact over a spherical surface while switching between pentamers and hexamers. Simulations with no energy difference between the two capsomer states yielded pronounced energy minima for structures containing 12, 32, 42 and 72 capsomers. These structures correspond precisely to \( T=1 \), \( T=3 \), \( T=4 \), and \( T=7 \) icosahedral capsids. When the energy difference between pentamers and hexamers is large, i.e. all the capsomers are in one of the two states, capsids with octahedral symmetry were found in addition to \( T=1 \) and \( T=3 \) icosahedral structures.

The assembly of protein capsids requires folding of the subunits and correct association of the complex. It is proposed that capsid assembly goes through a series of intermediate structures with lower assembly states, i.e. the capsomers, and that the structure with the largest interfaces will be the main intermediate and assemble first (13, 14). For example, bacteriophage HK97 forms a \( T=7 \) icosahedral capsid in which the pentameric and hexameric substructures have the largest
interface. In vitro, HK97 subunits first form pentamers and hexamers, which then assemble to capsids most efficiently when they are present in a ratio of 1:5 as in the final capsid structure (15). In an analysis of oligomeric proteins Levy et al. found that, in general, the largest interface in an oligomer is typically the evolutionary oldest, and when disassembly is induced, substructures containing the largest, evidently most stable interface were mainly detected (2).

1.3 Functions of protein capsids in nature

In addition to lipid membrane-bound organelles nature uses protein containers to achieve compartmentalization. Recent studies have made clear that prokaryotes are not devoid of intracellular organization and contain protein microcompartments that sequester specific metabolic processes. These protein organelles are highly symmetric polyhedral protein shells assembled from a small number of subunits reminiscent of viral capsids.

The most extensively studied bacterial microcompartment is the carboxysome (Fig. 1.2)(16). It enhances CO$_2$ fixation in cyanobacteria and some chemoautotrophs by colocalizing the two enzymes carbonic anhydrase and ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO) within its protein shell. Bicarbonate from the cytosol enters the carboxysome and gets converted by carbonic anhydrase to CO$_2$, which is then used as a substrate by RubisCO. Colocalization of the two enzymes results in substrate channeling which provides the catalytically inefficient RubisCO with high effective concentrations of CO$_2$. The outer protein shell not only ensures colocalization of the two enzymes it might also enhance CO$_2$ fixation by preventing CO$_2$ exit, resulting in even higher CO$_2$ concentrations, and by impeding entry of the competitive inhibitor O$_2$. The polyhedral protein shell has a diameter of about 100 nm and is mainly composed of proteins forming hexameric rings. In crystals these hexamers arrange into extended, tightly packed layers, which most likely form the facets of the polyhedral shell. Another carboxysomal protein was found to form pentamers and a structural model of the shell, consistent with electron microscopy data, was proposed in which these pentamers reside on the vertices of an icosahedral capsid (17). The necessary exchange of substrates and products
between the cytosol and the compartment interior might be accomplished by pores in the carboxysome shell, which are located in the center of the hexameric rings. The pore in one of the hexameric proteins has been found in an open and a closed form, which might allow gated transport of molecules across the protein shell (18).

Figure 1.2. Bacterial microcompartments. (A) Model of the carboxysome (17), hexamer (pdb: 3bn4, ccmk1), pentamer (pdb: 1qw7, ccmL). (B) Metabolic processes compartmentalized within the carboxysome, (C) the ethanolamine utilization and (D) the propanediol utilization microcompartments (adapted from reference (19)).

Other examples of bacterial microcompartments include ethanolamine utilization (Eut) and propanediol utilization (Pdu) microcompartments (Fig. 1.2) (19). They form very similar polyhedral structures to the carboxysome and mediate more complex metabolic pathways that are coenzyme B12 dependent. Ethanolamine and 1,2-propanediol are metabolized as carbon and energy sources by encapsulating the enzymes necessary for these processes within the protein shell. As in the carboxysome, a hexameric shell protein, forming a central pore that can open and close, has been identified in the Eut microcompartment (20). The open pore provides a potential route for transport of larger molecules such as cobalamin and...
other required cofactors across the protein shell. Sequestration of these pathways into microcompartments ensures optimum efficiency by substrate channeling. But it might also serve to protect the host cell from the toxic effects of the acetaldehyde and propionaldehyde or prevent loss of these volatile intermediates.

A common protein domain was identified in the main shell protein of diverse microcompartments that has become known as the bacterial microcompartment (BMC) domain. Comparative genomics approaches have revealed that open reading frames encoding proteins with BMC domains are found in over 400 of about 2000 microbial genomes (19). The wide distribution of the BMC domain suggests that more microcompartments with diverse functions exist and underscores the benefits of compartmentalization for living organisms.

No evidence for BCMs has been found in Archea, however. A different type of protein compartment, encapsulin, has been identified both in Archea and Bacteria (21). The protein shells of encapsulin are assembled from 60 identical monomers to form a T=1 icosahedral capsid with a diameter of 24 nm. Encapsulins sequester iron-dependant peroxidases or proteins related to ferritin, both involved in oxidative stress response. A short C-terminal extension on these enzymes that binds to a conserved binding site on the capsid interior drives their encapsulation. A nanocompartment of similar size is formed by lumazine synthase, which catalyzes the penultimate step in the biosynthesis of riboflavin (22). In *B. subtilis* lumazine synthase sequesters the next enzyme in the pathway, riboflavin synthase, within its icosahedral capsids, thereby improving riboflavin production by substrate channeling (23).

Vault particles are an example of protein containers in eukaryotes. They form barrel-shaped protein shells with maximum dimensions of 42 x 75 nm and are highly abundant with $10^4$-$10^5$ particles per cell in the cytosol of most eukaryotes, including humans (24). Although many different functions have been proposed for these structure, including roles in intercellular and nucleocytoplasmic transport, multidrug resistance, and cell growth and proliferation, the precise cellular function remains unknown.

Among prominent examples of natural protein capsids is the iron storage protein ferritin, which is found in all domains of life (25). It forms an octahedral protein capsid of 24 identical subunits with a diameter of 12 nm and catalyzes the
mineralization of iron inside its cavity as a storage mechanism to maintain iron homeostasis. Gated pores control the flow of iron across the protein shell. Additional well-characterized examples of protein containers in nature include viruses (26) and chaperones (27) that assist protein folding.

1.4 Engineering protein capsids for bionanotechnology

Encapsulation provides control over the sequestered guests and their reactions by regulating entry and exit of potential substrate molecules. Colocalization of two or more catalysts in a protein chamber can enhance catalytic efficiency by substrate channeling. Sequestration of a protein in the cell is a simple mechanism to prevent side reactions and inhibition or to protect the surrounding environment from the toxic effects of a molecule. The container interior can offer a tailored microenvironment, different from the surrounding milieu, which favors the desired process. Furthermore, encapsulated proteins, intermediates, and products are protected by the capsid shell. The natural protein container ferritin demonstrates how protein capsids are able to render sequestered material (like solid iron oxide) biocompatible and soluble. Enclosing proteins in the confined space of capsids is believed to stabilize folded over extended unfolded states and thereby improve folding (28, 29). This strategy is utilized by chaperones (30). Encapsulation also increases thermal stability and resistance to cellular proteases (31).

Nature has demonstrated how the unique and beneficial properties of compartmentalization can be harnessed. There is currently considerable interest in mimicking nature and exploiting encapsulation for the development of systems for diverse applications such as gene and drug delivery, bio- and medical imaging, catalysis, and controlled synthesis of novel materials. Towards this goal, many different types of material have been used to build compartments including polymers, dendrimers, silica-based particles, liposomes (32), supramolecular organic capsules (33), and cages with natural building blocks like DNA polyhedra (34, 35) and protein capsids (36, 37). Protein containers provide particularly attractive scaffolds since they spontaneous self-assemble into highly organized and monodisperse structures of defined size. Depending on the system, the dimensions
of protein capsids can range from 10 to 500 nm (Fig. 1.3). Moreover, such structures are inherently biocompatible and they can be easily modified by both genetic and chemical methods.

Figure 1.3. Natural protein containers. Structures of protein capsids based on X-ray crystallography and electron microscopy. Images were made using Chimera (38) and Viperdb (39). PBCV-1 is paramecium bursaria chlorella virus type 1, pdb: 1m4x. HK97 is bacteriophage HK97, pdb: 1ohg. CCMV is cowpea chlorotic mottle virus, pdb: 1za7. MS2 is bacteriophage MS2, pdb: 2ms2. TMV is tobacco mosaic virus, pdb: 2om3. LS is lumazine synthase from Aquifex aeolicus, pdb: 1hqk. sHSP is small heat shock protein from Methanococcus jannaschii, pdb: 1shs. Ferritin is wild-type human ferritin, pdb: 3ajo. Dps is DNA binding protein-like from Listeria innocua, pdb: 2bij.

Encapsulation of inorganic and polymeric materials

Ferritin was one of the first protein capsids to be used for size-constrained synthesis of monodisperse inorganic nanoparticles (40-42). Taking advantage of its natural ability to sequester and mineralize iron, iron sulfide and iron oxide and
subsequently many other inorganic nanoparticles have been synthesized inside the ferritin cavity in biomimetic strategies (36). The ferritin cage not only controls the size of the resulting nanoparticle, it can also be used for the formation of new materials with controlled composition and properties. In one example, the preparation of mixed cobalt and iron oxide nanoparticles in ferritin yielded a composite material with magnetic properties, which could be tailored by varying the synthesis conditions and the cobalt/iron ratio (43). Controlling the properties of such nanomaterials is an important ability for the development of new devices.

Engineered ferritin cages have been used for the synthesis of metallic silver and gold nanoparticles. In one case, a silver binding peptide that had been identified by phage display was introduced at the inner surface of ferritin. It promoted the selective synthesis of silver nanoparticles only within the capsid (44). In another example, the synthesis of gold and silver nanoparticles within ferritin capsids was improved by removing solvent exposed cysteines and histidines on the outside of the container and adding cysteines to the interior surface of the capsid (45).

Ferritin is not the only protein capsid that can serve as a template for the synthesis of inorganic and polymeric materials (36, 46). A member of the ferritin protein family, the Dps proteins, forms smaller cages with an outer diameter of only 9 nm (compared to 12 nm for ferritin). It has also been used for the synthesis nanoparticles like ferrimagnetic iron oxide (47). Other protein capsids that have been employed include cages from viruses like the Cowpea chlorotic mottle virus (CCMV) which is about 28 nm in diameter, the small heat shock protein (sHsp), chaperonins, and lumazine synthase. The tobacco mosaic virus (TMV) with its tubular structure (300 x 18 nm) has been used for the synthesis of Ni and Co nanowires that have uniform widths of 3 nm wide and lengths of up to 600 nm (48). Protein capsids further provide the ability to arrange nanoparticles in two-dimensional layers (49, 50). This capability was demonstrated by the synthesis of Ni-Pd and Co-Pd nanoparticles within engineered chaperonins that had been assembled in 2D arrays (50).

An alternative approach to create protein-encapsulated nanoparticles is to assemble the protein capsids around preformed nanoparticles. Gold nanoparticles functionalized with citrate, carboxylate-terminated tetraethylene glycol or RNA could be encapsulated in this way into viral capsids (51-53). The approach was
extended to the encapsulation of polyethylene glycolated CdSe/ZnS semiconductor quantum dots which enhanced their stability (54). Chaperones have been used to encapsulate CdS semiconductor nanoparticles, also increasing their thermal and chemical stability in aqueous media. Importantly, the nanoparticles could be released from the chaperone cavities upon ATP hydrolysis, exploiting the natural mechanism by which proteins are released after folding is complete (55). Besides stabilizing and arranging nanoparticles, the protein shell also makes them biocompatible. As a consequence, protein-caged gold and quantum dot nanoparticles with useful optical and magnetic properties have considerable potential for utilization as biomarkers or -sensors.

As these examples show, protein cages provide reaction vessels for the controlled synthesis of nanoparticles with defined dimensions and a very narrow size distribution. The resulting nanomaterials can possess special properties not accessible by conventional syntheses. Due to the intrinsic organization of protein capsids they can be further assembled into higher ordered two- and three-dimensional arrays (56). Protein capsids therefore are important elements in the development of new nanomaterials for applications as biomarkers, catalysts, and in electronic, magnetic and bioresponsive devices.

**Bioimaging and drug delivery**

Encapsulating biologically or medicinally relevant molecules imparts protein capsids with great potential for applications as diverse as in vivo imaging and gene and drug delivery. The protein shell improves biocompatibility, solubility and stability, and reduces toxicity effects associated with the encapsulated compounds. Further, it can be easily modified by genetic and chemical methods to target specific cells or tissues, to promote cell entry, and to lower an immune response. Although the intrinsic ability of some viruses to enter specific cells can be exploited, their properties often fail to meet therapeutic needs (57, 58). Therefore, a range of peptides, aptamers, antibodies, carbohydrates, polymers and other molecules (often derived from natural sources) have been developed to direct protein capsids to specific cell types or to support cell entry.

Protein cages filled with optical or magnetic nanoparticles or metal complexes, chromophores or fluorescent dyes are proving useful in medical imaging. For
example, they have been used to examine the vascular system and to visualize cancer cells. In biological imaging they have found application as microscopic probes e.g. to track viruses and other biologically important processes (59, 60). Interesting potential applications in magnetic resonance imaging (MRI) have been demonstrated by caging magnetic resonance contrast agents in protein capsids to give complexes with higher relaxivity and improved water solubility. One approach towards targeted imaging has exploited iron oxide-containing or fluorescently labeled ferritins displaying a tumor targeting peptide, which had been identified by phage display, to bind to cancer cells in vitro (61).

Examples of encapsulating therapeutic molecules include the attachment of the antitumor agent doxorubicin to the interior of the small heat shock protein and its selective release under acidic conditions, such as those found in endosomes (62). The poor water-solubility of the chemotherapeutic drug taxol necessitates its administration with a toxic detergent cocktail over a long period. Encapsulation into protein cages might be an alternative approach with less harmful side effects. To that end, taxol was conjugated to the interior of bacteriophage MS2 capsids; the taxol-modified MS2 was shown to decrease the cell viability of breast cancer cells in vitro (63). Caging drugs by protein shells can also circumvent existing resistance as shown in a study of methotrexate (MTX)-loaded polyomavirus capsids (64). GFP and MTX were encapsulated by covalent linkage to a domain of an inner core protein in the lumen of the virus capsids. The caged GFP and MTX were internalized into leukemia T cells and MTX-transport resistant cells, allowing the cytostatic effects of MTX to be manifest.

Gene therapy is currently one of the most promising approaches for treatment of cancer and genetic diseases. Protein capsids, particularly those based on viral shells, are attractive as possible delivery vehicles (57). Capsids derived from hepatitis B virus (L particles) loaded with the genes encoding the human blood clotting factor IX or GFP, or with a fluorescent dye were efficiently delivered to the target cells both in cell culture and in a mouse xenograft model (65). After delivery, factor IX was produced in the xenograft model at levels relevant for the treatment of hemophilia. The cell-targeting specificity of the L particles could be altered by substituting their natural recognition motif with human epidermal growth factor.
A crucial factor for gene and drug delivery is cell entry and membrane penetration. In a study with vault particles, which are naturally present in humans and therefore neither immunogenic nor toxic, it was shown that a viral domain with membrane lytic activity could be incorporated into the complex in functional form (66). The modified vault particles containing a soluble ribotoxin or plasmids encoding GFP were found to be internalized more efficiently than unmodified vaults.

Despite numerous advances and promising results in engineering viral and other protein capsids for gene and drug delivery, such systems have not been used for clinical treatments yet. Further engineering based on increased knowledge about the involved physiological processes can be expected to yield optimized protein delivery vehicles that can be used for the treatment of many diseases. Implementation of powerful screening and selection techniques may speed up this process (57).

**Catalysis**

As impressively demonstrated by natural examples of micro- and nanocompartments, compartmentalization provides significant advantages with respect to the control and efficiency of catalysis.

Nanoparticles inside protein capsids can act as catalytic sites. For example, small well-defined Pt clusters, which were synthesized in the cavity of the small heat shock protein, efficiently catalyze the reduction of H⁺ to H₂ in a light driven reaction (67, 68). The activity of these encapsulated nanoparticles is higher than in other Pt nanoparticles, presumably because sequestration in the protein cage stabilizes the Pt catalyst and prevents agglomeration. The pores of protein cages also enable control over substrate access to such encapsulated catalysts as shown by the size selective olefin hydrogenation of a ferritin-caged Pd cluster (69).

Some reactions involving multiple steps require the precise positioning of the different catalytic centers. For example, relative distances and orientation are crucial for efficient energy and electron transfer reactions. Natural photosynthetic systems achieve such organization by self-assembling complexes of membrane-bound proteins. The self-assembly of protein capsids into highly organized structures has been exploited to create an artificial photocatalytic system (70).
Fluorescent donors and a porphyrin acceptor were attached to specific residues of the bacteriophage MS2 resulting in defined positioning of the donors inside the capsid and the porphyrins on the outside. This arrangement enabled efficient energy transfer from the donors, stimulated by light, to the porphyrins. In this way, the porphyrins could be sensitized at previously inaccessible wavelengths, enabling catalysis of photoinduced reduction of methyl viologen at multiple donor-dependant wavelengths.

More recently, researchers have begun to encapsulate protein catalysts. In one example, CCMV was assembled around horseradish peroxidase (HRP) molecules at concentrations where not more than one HRP enzyme is present statistically in the capsid (71). The resulting complex allowed investigations of enzyme activity at the single-molecule level and showed that diffusion of substrates and products across the capsid can be controlled by the pH-dependent size of the pores in the CCMV shells. CCMV capsids were also used to encapsulate a precise number of multiple lipase enzymes (72). The sequestered lipases showed higher activity than the non-encapsulated enzymes free in solution. In another example, encapsulation of enzymes in bacteriophage Qβ virus-like particles was promoted by an RNA aptamer sequence that links the tagged enzymes to a viral RNA packaging hairpin, which in turn binds to the capsid interior (73). The activity and stability of the sequestered enzymes was compared to the free enzymes. Caged enzymes were highly active, but exhibited a somewhat lower catalytic efficiency than the free enzymes. Nevertheless, the protein shell stabilized them against thermal denaturation, proteolytic degradation and inactivation by adsorption to hydrophobic surfaces.

The stabilizing and protective effect of sequestration by protein capsids can be utilized for the production of labile, aggregation-prone or toxic proteins and enzymes. This was demonstrated by the soluble expression of difficult-to-produce proteins upon fusion to the central cavity of a chaperonin (74). Encapsulation further provides the opportunity to study the functional and structural properties of proteins in confined spaces and crowded environments, conditions that more resemble their natural setting within the cell than in dilute solution.

The possibility of selectively encapsulating active enzymes paves the way towards engineering of more sophisticated systems. Upon compartmentalization of
multiple enzymes artificial microcompartments might be generated that perform whole metabolic sequences.

What drives encapsulation?

Encapsulation of molecules into protein containers can be achieved by chemical attachment of the desired guest using selective bioconjugation reactions (63, 70). Often, strategies relying on natural mechanisms and recognition elements are employed. For example, specific incorporation of non-natural proteins or peptides into viruses or virus-like particles has been achieved by their covalent fusion to capsid shell proteins or domains of proteins that naturally attach to the inner capsid surface (64, 75). Noncovalent binding of guests to the capsid interior is possible by artificially incorporating oligomerization domains such as coiled-coil motifs. The utility of this latter approach was shown in a study in which encapsulation was realized by attaching one part of a heterodimeric coiled-coil to GFP and the other part the to the CCMV capsid protein (76). A heterodimeric motif was chosen to prevent self-dimerization. Capsid proteins bound to GFP were then reassembled to form highly loaded containers. Moreover, the amount of GFP loaded per capsid could be controlled by adding different amounts of wild-type capsid protein to the assembly reaction.

To drive encapsulation, the intended guest can also be functionalized with viral RNA encapsidation sequences that bind to the capsid interior (53, 77). The RNA has to be conjugated to the guest in vitro, which can sometimes be limiting. The aptamer strategy described above circumvents this problem by using an mRNA with two binding domains (an aptamer that binds to the protein tag and a viral packaging sequence that binds to the capsid interior) to link the guests to the capsid interior. The natural metal binding sites in ferritin and viruses have similarly been utilized to bind inorganic material (36, 61, 78). In a complementary strategy encapsulation can also be driven by simple electrostatic interactions. This strategy has been mainly used with highly charged inorganic and organic guest molecules for the synthesis of nanomaterials (79, 80), but can also be exploited for proteins as shown by an encapsulation system developed in our laboratory and extended in this thesis (81).
1.5 Lumazine synthase

Lumazine synthase (LS) catalyzes the penultimate step in the biosynthesis of riboflavin (vitamin B₂). It can be found in bacteria, archaea, plants, and fungi. Animals, however, depend on exogenous sources of riboflavin. More specifically, LS catalyzes the condensation of 5-amino-6-ribitylamino-2,4(1H,3H)-pyrimidinedione with 3,4-dihydroxybutanone 4-phosphate to form 6,7-dimethyl-8-ribityllumazine (Fig. 1.4) which is then converted by riboflavin synthase to riboflavin.

![Figure 1.4. The reactions catalyzed by lumazine synthase and riboflavin synthase.](image)

Crystal structures of LS from different organisms have revealed the occurrence of three oligomerization states (82-85). In fungi and some bacteria, LS forms homopentamers, LS from *Brucella abortus* assembles as D₅-symmetric dimers of pentamers, whereas these pentameric building blocks assemble further to icosahedral capsids in *Bacillus subtilis, Bacillus anthracis, Escherichia coli, Spinacia oleracea* and *Aquifex aeolicus*. Notably, the pentameric building blocks in all these structures are very similar (Fig. 1.5). Icosahedral LS have been crystallized as T=1 capsids built from 60 identical subunit, arranged as 12mers of pentamers, with an outer diameter of about 160 Å (84). However, depending on the choice of buffer, pH and presence of certain ligands, they can adopt multiple assembly states (86). In addition to T=1 states, capsids with T=3 symmetry and a diameter of around 29
nm, and even larger structures consistent with either an expanded T=3 or a T=4 state have been observed. In *Bacillus subtilis* the icosahedral LS naturally encapsulates a trimer of riboflavin synthase in its cavity improving the efficiency of riboflavin biosynthesis by substrate channeling (23).

![Figure 1.5](image.png)

**Figure 1.5.** Structures of lumazine synthases. Structure of the icosahedral T=1 capsids of *A. aeolicus* LS (left) (84) and the pentameric structure of *S. cerevisae* LS (right) (82).

The tunnels in the center of the pentameric ring have been proposed to allow diffusion of substrates into the interior of the capsid, where the active sites are located at subunit interfaces. The width of the opening (Cα to Cα) is ∼16 Å, which appears to be too narrow for passage of lumazine and riboflavin. This observation suggests that capsid dynamics is probably quite important. In fact, a mutant of *Aquifex aeolicus* LS formed T = 3 capsids with expanded pentamers containing a widened channel. A model for catalysis was suggested, in which LS cycles between an open inactive conformation, allowing binding of substrates and release of products, and a closed active form, which performs catalysis (87).

The icosahedral capsids of lumazine synthases present an attractive alternative to ferritins, viruses, and heat shock proteins for the engineering of protein containers with new functions. *B. subtilis* LS has been shown to be a viable template
for the size-constrained synthesis of iron oxide nanoparticles (88). In our laboratory a protein encapsulation system based on LS from *Aquifex aeolicus* (AaLS) has been developed (81). AaLS originates from a hyperthermophilic bacterium growing in hot springs with maximum temperatures of up to 95 °C and is therefore extremely stable and can be expressed in *E. coli* in high yields. To drive encapsulation into the AaLS capsids electrostatic interactions between host and guest were implemented (Fig. 1.6). Four glutamates per AaLS monomer were introduced at the inner surface of the capsid, resulting in a variant (AaLS-neg) with 240 or 720 additional negative charges per T = 1 or T = 3 capsid. Appending a positively charged deca-arginine tag to the C-terminus of green fluorescent protein (GFP-R10) and coproduction with AaLS-neg in *E. coli* led to association of three to four GFP-R10 molecules per T = 3 capsid. AaLS should constitute a robust system for encapsulation of a variety of proteins.

![Figure 1.6](image1.png)

**Figure 1.6.** Schematic representation of the AaLS encapsulation system.

### 1.6 Aims of this thesis

The aims of this thesis were to optimize the rationally designed AaLS encapsulation system and extend its functions. Specifically, the goals included sequestration of other guests, development of alternative loading methods, and creation of function in the LS channels. We anticipated that caging a toxic protein might be a mechanism for protecting the host cell, which should lead to a selectable advantage that could be exploited to improve encapsulation by directed evolution.

Chapter 2 describes the successful directed evolution of AaLS capsids. Using sequestration of the toxic R10-tagged HIV protease as the basis for the development
of a selection system, a variant (AaLS-13) with higher loading capacity was obtained which protects the host cells against increased intracellular concentrations of HIV protease.

Detailed characterization of the encapsulation system is outlined in Chapter 3. The specificity of guest sequestration in vivo, the encapsulation of alternative guests, and the stability of the complexes are examined. Loading with diverse positively charged proteins is possible and results in stable host-guest complexes.

The loading of AaLS-13 with a positively charged protein in vitro is shown in Chapter 4. Controlled disassembly of the “empty” capsids and reassembly around a suitable guest resulted in highly loaded AaLS-13 capsids. Establishing conditions that ensure reproducible and efficient reassembly and loading is important toward providing useful routes to well defined complexes for practical applications.

The tunnels in the pentameric building blocks of LS connect the capsid interior to the bulk surrounding. Chapter 5 describes attempts to create de novo catalytic activity in the tunnels of the pentameric yeast LS. A simple symmetry-based rational design yielded an active variant for a simple retro-aldol reaction.

The results obtained in this thesis demonstrate that protein capsids from AaLS are valuable scaffolds for the engineering of encapsulation systems. They lay the foundation for many exciting applications, including the construction of multienzyme complexes, biosensors, nanomaterials and novel delivery vehicles.
2 DIRECTED EVOLUTION OF A PROTEIN CONTAINER

2.1 Introduction

Compartmentalization is a hallmark of life on earth. By sequestering genes and gene products from the environment, genotype becomes linked to phenotype, enabling Darwinian evolution. By restricting reaction pathways to specific compartments, otherwise incompatible metabolic processes can also proceed simultaneously within the cell. For such tasks, Nature generally makes use of large membrane-bounded structures, either whole cells or intracellular organelles. For many specialized biological functions, however, smaller shell-forming proteins are the ideal molecular container. For example, protein compartments play essential roles in protein folding [chaperones (27)], storage and transport of minerals [ferritin(25)], transfer of nucleic acids [viruses (26)], and regulation of enzyme catalysis [pyruvate dehydrogenase (89), fatty acid synthase (90), the carboxysome (16), and the ethanolamine utilization microcompartment (20)].

Inspired by nature, nanotechnologists have begun to exploit the properties of natural protein containers, particularly capsid-forming viruses, toward diverse applications including gene therapy (57), drug delivery (62, 64, 65), bioimaging (61, 91, 92), catalysis (37), and controlled synthesis of novel inorganic and polymeric materials (36, 43, 47, 48, 50, 55, 93). Polyhedral protein compartments are attractive in this context due to their biocompatibility, spontaneous self-assembly into highly organized structures of defined size, and easy modification by genetic methods. Here, we take advantage of these features to engineer an artificial protein container capable of spontaneously encapsulating a toxic protease in the cytoplasm of E. coli.
2.2 Results

Lumazine synthase from *Aquifex aeolicus* (AaLS) (84) forms icosahedral capsids consisting of 60 (T = 1 state) or 180 (T = 3 state) identical subunits that catalyze the penultimate step in riboflavin biosynthesis. Outside the cell, the hollow spheres formed by AaLS have been utilized as templates for crystallization of inorganic nanoparticles (88). The AaLS capsid has also been redesigned to bind other guest molecules unrelated to its normal biochemical function by employing engineered electrostatic interactions between host and guest to drive encapsulation (81). Introduction of four glutamates per monomer (R83E, T86E, T120E, and Q123E), plus a C-terminal His$_6$-tag for purification, afforded AaLS-neg, a variant possessing 240 to 720 additional negative charges on the luminal surface of its T = 1 and T = 3 capsids, respectively. The negatively supercharged capsid preferentially encapsulates cargo proteins containing a complementary positively charged peptide tag. Thus, coproduction with green fluorescent protein bearing a C-terminal deca-arginine sequence (GFP-R$_{10}$) affords a cage complex containing three to four GFP molecules per T = 3 capsid.

2.2.1 Encapsulation of HIV protease by AaLS-neg

As in vivo association of AaLS-neg and guest relies on the appended R$_{10}$ sequence, packaging of any appropriately tagged molecule should be possible. For example, we anticipated that attaching the R$_{10}$ tag to the C-terminus of an enzyme that is toxic to *E. coli* when produced cytoplasmically, and coproduction with the capsid AaLS-neg would lead to selective encapsulation of the toxic protein in vivo and thereby provide the host cell with a selectable advantage. We initially chose barnase, a ribonuclease, as the toxic protein. However, wild type barnase and variants with diminished activity (E73Y, E73A, K27A) were too toxic for the *E. coli* host cell.

HIV protease is another example of an enzyme that is toxic to *E. coli* when produced cytoplasmically. Appending the R$_{10}$ tag to the C-terminus of the dimeric HIV protease (94) and coproduction with AaLS-neg should lead to selective
encapsulation of the protease in vivo (Fig. 2.1A). As a result, bacterial proteins would be shielded from proteolytic degradation, providing the host with a selectable advantage (Fig. 2.1B). In contrast, wild-type AaLS would not be expected to associate with the tagged protease and should therefore be unable to counter its deleterious effects.

Figure 2.1. Encapsulation of toxic proteins. (A) An engineered variant of lumazine synthase from Aquifex aeolicus (grey) forms icosahedral capsids that spontaneously assemble around HIV protease bearing a deca-arginine tag (blue). The four residues per monomer that were mutated to glutamates (red) point into the luminal space of the assembled capsid, where they can interact electrostatically with the positively charged guest molecule. (B) While production of the tagged protease is lethal to the host (left), encapsulation by the engineered lumazine synthase rescues the cells (right). The selectable growth advantage was exploited for directed evolution to enhance the efficiency of encapsulation.

To control the biosynthesis of the toxic enzyme, its gene was expressed from the tightly regulatable tetracycline promoter (95). In the absence of tetracycline, HIV protease is not produced, so cells exhibit normal growth independent of the co-produced capsid variant (Fig. 2.2A). However, moderate induction of the enzyme with 450 ng/ml tetracycline in liquid culture greatly impairs cell viability. As expected, this phenotype is not altered by the presence of wild-type AaLS, but co-production of AaLS-neg partially restores cell growth (Fig. 2.2B). However, the ability of AaLS-neg to rescue cell growth does not withstand a further increase in the inducer concentration to 700 ng/ml (Fig. 2.2C). The growth advantage accruing
to AaLS-neg compared to AaLS-wt in the presence of HIV protease-R_{10} over a range of induction levels is summarized in Fig. 2.2D.

**Figure. 2.2.** Protective effect of AaLS-neg on cells expressing HIV protease-R_{10}. (A), (B), (C) and (D) Growth curves of cells transformed with plasmids encoding HIV protease-R_{10} and either AaLS-wt (green), or AaLS-neg (blue). Cultures were grown in liquid LB media at 30 °C. Production of AaLS variants was induced with 100 µM salicylate. Expression of the protease gene was controlled by the tightly regulatable tetracycline promoter. In the absence of tetracycline (A) the protease is not produced, while addition of 450 ng/ml (B) or 700 ng/ml (C) tetracycline leads to increasing intracellular concentrations of the toxic enzyme. (D) OD_{600} values of the growth curves in (A)-(C) after 26 h were plotted against the tetracycline concentrations used for HIV protease-R_{10} induction.

Growth tests with HIV protease lacking the R_{10} tag show that the tag makes an important contribution to successful rescue of cell growth by AaLS-neg (Fig. 2.3). In this case, AaLS-neg only provides a small advantage over AaLS-wt in a very narrow range of tetracycline concentrations (Fig. 2.3D). In the absence of the positively charged tag, host-guest assembly is apparently very inefficient.

Cell growth on solid media confirms the results of the liquid growth tests (Fig. 2.4). The growth advantage achieved by coproducing the HIV protease with AaLS-neg compared to AaLS-wt and is readily apparent on solid media, showing that the R_{10} tag is necessary for efficient rescue.
Figure 2.3. Growth of cells producing AaLS-wt or AaLS-neg and HIV protease lacking the R₁₀-tag. (A), (B), (C) and (D) Growth curves of cells transformed with plasmids encoding HIV protease and either AaLS-wt (green), or AaLS-neg (blue) as described in Fig. 2.2, but at tetracycline concentrations of 0 mg/ml (A), 100 ng/ml (B) or 133 ng/ml to induce protease expression (C). (D) OD₆₀₀ values of the growth curves in (A)-(C) after 26 h were plotted against the tetracycline concentrations used for HIV protease-R₁₀ induction.

Figure 2.4. Cells growth on solid media. (A) and (B) Complementation frequencies of cells transformed with plasmids encoding AaLS-wt (green) or AaLS-neg (blue), and HIV protease either R₁₀-tagged (A) or untagged (B). Cells were grown on LB-agar plates at 30°C. Production of AaLS variants was induced with 100 μM salicylate. Expression of the protease gene was induced at different levels by tetracycline.
Western blot analysis of capsids isolated by Ni$^{2+}$-affinity chromatography established that a significant amount of the R$_{10}$-tagged protease copurifies with AaLS-neg (Fig. 2.5) but not with wild-type AaLS, suggesting that the growth advantage provided by the engineered capsid derives from its intimate association with the enzyme. As seen with GFP (81) and in accordance with the growth tests, this interaction is dependent on both design features, the negatively charged inner surface of the capsid and the positively charged R$_{10}$ tag (Fig. 2.5).

**Figure 2.5.** Encapsulation of HIV protease. SDS-PAGE (left) and Western blot (right) analysis of purified AaLS-wt and AaLS-neg capsids coproduced with HIV protease, either R$_{10}$-tagged or untagged. Capsid protein was visualized by SDS-PAGE, while the relative amounts of copurifying HIV protease were determined by Western blotting using a commercial murine anti-HIV protease antibody. The small amount of untagged protease observed in the HIV protease-R$_{10}$ samples may result from autoproteolysis, which would imply that the encapsulated enzyme is catalytically active, or from cleavage by cellular proteases prior to encapsulation.

### 2.2.2 Directed evolution

Host-guest pairing proffers a selectable phenotype that can be exploited to optimize this rationally designed encapsulation system by directed evolution. Model selections established a selection protocol (Fig 2.6) and demonstrated successful enrichment of cells coproducing the AaLS-neg capsid over those coproducing the wild type capsid with the HIV protease-R$_{10}$ of up to 200-fold in one selection round. GFP and BFP served as reporter protein for the cells containing AaLS-neg and AaLS-wt, respectively, and their genes were cloned biscistronically into the capsid-containing vector. These model selections showed that transformation of the HIV protease-R$_{10}$ containing plasmid as a last step before the selection is crucial to minimize artifacts arising from mutations in the HIV protease gene or the plasmid.
For the directed evolution of improved capsid variants, AaLS-neg served as the starting point. Using the gene encoding AaLS-neg as a template, a large library of mutants was created by error-prone PCR. Plasmids encoding the capsid library and the R_{10}-tagged protease, respectively, were sequentially electroporated into *E. coli* XL1-Blue cells. The transformants were then grown in liquid culture at 30 °C. Selection pressure was applied by adding tetracycline to induce protease production. Because cells containing the starting AaLS-neg capsid variant did not grow at or above 600 ng/ml tetracycline, the pool of transformed cells was split and selections were performed in parallel at tetracycline concentrations ranging from 600 to 1200 ng/ml. Plasmids encoding permissive capsid variants were isolated from the survivors, pooled, further diversified by DNA shuffling to combine beneficial mutations, and then used as input for a second round of selection. A third round of diversification by DNA shuffling and selection, and a fourth round of selection without further diversification, were carried out. For all selection rounds library sizes ranged between 3 \times 10^6 and 3 \times 10^7 transformants (Table 2.7). During each round of selection, cultures containing the library grew more rapidly than in the previous round, indicating accumulation of beneficial changes (Fig. 2.8).
Table 2.7. Library statistics. The number of transformants is given for the initial electroporation of the plasmids encoding the capsid library and for the subsequent electroporation of the plasmid encoding HIV protease-R_{10}. For each round, sequences of AaLS variants were analyzed before and after selection at both the DNA and the amino acid level. The number of mutations that were observed in at least two clones (defined as common mutations) increased after each round of selection.

<table>
<thead>
<tr>
<th>Library transformsants</th>
<th>Round I</th>
<th>Round II</th>
<th>Round III</th>
<th>Round IV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$5 \times 10^6$</td>
<td>$3 \times 10^6$</td>
<td>$3 \times 10^7$</td>
<td>$8 \times 10^6$</td>
</tr>
<tr>
<td>HIV protease-R_{10} transformsants</td>
<td>$1 \times 10^7$</td>
<td>$4 \times 10^7$</td>
<td>$4 \times 10^6$</td>
<td>$2 \times 10^6$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DNA level:</th>
<th>unselected</th>
<th>selected</th>
<th>unselected</th>
<th>selected</th>
<th>unselected</th>
<th>selected</th>
<th>unselected</th>
<th>selected</th>
</tr>
</thead>
<tbody>
<tr>
<td># sequences considered</td>
<td>21</td>
<td>28</td>
<td>8</td>
<td>18</td>
<td>4</td>
<td>23</td>
<td>6</td>
<td>24</td>
</tr>
<tr>
<td>mutations per gene (range)</td>
<td>1-7</td>
<td>1-6</td>
<td>4-9</td>
<td>4-10</td>
<td>6-13</td>
<td>3-14</td>
<td>8-10</td>
<td>6-14</td>
</tr>
<tr>
<td>mutations per gene (average)</td>
<td>3 ± 0.6 %</td>
<td>2.8</td>
<td>6.3</td>
<td>6.8</td>
<td>8.5</td>
<td>8.2</td>
<td>7.7</td>
<td>9.3</td>
</tr>
<tr>
<td>standard deviation</td>
<td>1.5</td>
<td>1.3</td>
<td>1.9</td>
<td>2.0</td>
<td>3.1</td>
<td>2.5</td>
<td>3.9</td>
<td>2.2</td>
</tr>
<tr>
<td>frame shifts</td>
<td>4</td>
<td>0</td>
<td>2</td>
<td>6</td>
<td>3</td>
<td>7</td>
<td>1</td>
<td>9</td>
</tr>
</tbody>
</table>

Amino acid level:

<table>
<thead>
<tr>
<th># sequences considered</th>
<th>selected</th>
<th>common mutations</th>
<th>selected</th>
<th>common mutations</th>
<th>selected</th>
<th>common mutations</th>
<th>selected</th>
<th>common mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>mutation per protein (range)</td>
<td>0-5</td>
<td>0-2</td>
<td>3-8</td>
<td>1-3</td>
<td>1-8</td>
<td>1-6</td>
<td>4-9</td>
<td>3-7</td>
</tr>
<tr>
<td>mutation per protein (average)</td>
<td>1.9</td>
<td>0.8</td>
<td>5.2</td>
<td>1.9</td>
<td>3.8</td>
<td>2.1</td>
<td>7.1</td>
<td>4.4</td>
</tr>
<tr>
<td>standard deviation</td>
<td>1.4</td>
<td>0.8</td>
<td>1.5</td>
<td>0.8</td>
<td>1.9</td>
<td>1.5</td>
<td>1.8</td>
<td>1.2</td>
</tr>
</tbody>
</table>

Figure 2.8. Cell growth during selection rounds. 24 h after start of the selection by induction of HIV protease-R_{10} with tetracycline (1200 ng/ml), the OD values of the cultures were measured at 600 nm to assess the relative growth for cells containing capsids AaLS-wt (green), AaLS-neg (blue) and an AaLS-library (red).
Figure 2.9. Amino acid sequence alignment of 24 clones selected in the fourth round of selection. The first 14 sequences in the alignment are those of the full-length variants, with intact His_{6}-tags, that were produced and screened for encapsulation efficiency by ELISA. As noted in the text, AaLS-13 showed the greatest binding capacity and was characterized in detail. The vector of clone 23 contains two AaLS gene inserts, the second of which is not in frame. AaLS-21 has a deletion of one nucleotide at position 412 leading to a frame shift. AaLS-3, 11, 12, 15, 16, 22, 24 all have a deletion of one nucleotide at position 364. Positions corresponding to ambiguous or unreadable DNA sequencing data are marked by x. Stop codons are indicated by an asterisk (*).

DNA sequencing of representative clones after each round revealed the enrichment of specific mutations. Although no unique consensus sequence emerged, the set of 24 selected clones sequenced after round four include ten variants with frame shifts (Fig. 2.9); this drastic structural change signals a novel mechanism of cell rescue compared to AaLS-neg, such as trapping of the protease in inclusion bodies. Because the remaining 14 full-length variants were more likely to form
conventional capsids capable of complexing the protease via encapsulation, they were isolated by Ni$^{2+}$-affinity chromatography and analyzed by ELISA for copurifying HIV protease-R10 (Fig. 2.10). Two of these variants showed a significantly higher amount of associated HIV protease, which was confirmed by Western blot (Fig. 2.10). The best binder, AaLS-13, was characterized in detail.

Figure 2.10. Encapsulation capabilities of selected variants after evolution round IV. The amount of HIV protease-R10 that copurifies with the selected capsid variants upon their coproduction was detected by ELISA (left) and Western blot (right) using an anti-HIV protease antibody. In the majority of the variants, no associated protease could be detected by ELISA. These capsid variants all displayed low solubility in vivo, suggesting that cell rescue might occur by coprecipitation of the protease with the capsid as inclusion bodies. Interestingly, all these variants share two mutations, N102I and E106G. Capsid variants with frame shifts could conceivably act by the same mechanism. For ELISA screening, capsids were produced using the T7 promoter system and induction with IPTG, whereas the salicylate promoter system and induction with salicylate was used in the selections. The difference in expression levels obtained with these systems might also lead to differences in the concentration of soluble protein in the cell, so other mechanisms of cell rescue during the selection experiments cannot be ruled out.

2.2.3 Characterization of the evolved variant AaLS-13

Variant AaLS-13 binds five to ten-times more protease than AaLS-neg, based on Western blot analysis (Fig. 2.11A). Quantification of the absolute number of HIV protease molecules encapsulated per AaLS-13 capsid by Western blot indicates that approximately seven HIV protease dimers are bound per 180mer cage (Fig. 2.11B).
Figure 2.11. Encapsulation of HIV protease-R10 by AaLS-13. (A) The increased loading capacity of AaLS-13 compared to AaLS-neg was quantified by SDS-PAGE (top) and Western blot (bottom) using a commercial murine anti-HIV protease antibody. The relative amount of HIV protease-R10 bound by AaLS-13 was estimated by loading serial dilutions of this sample, as indicated above each lane, and comparing the band intensities with that of the sample containing AaLS-neg. (B) The number of HIV protease molecules per AaLS-13 capsid was determined by Western blot of the complex, loaded at the capsid concentrations indicated above each lane, and comparison with known amounts of wild type HIV protease. (C) and (D) Production of soluble HIV protease-R10 in the presence and absence of AaLS capsids. (C) Production using typical selection conditions. E. coli X1-Blue cells transformed with plasmids encoding the relevant capsid variant and the tagged HIV protease-R10 were grown to an OD₆₀₀ of 0.5. Capsid and protease production were then induced with 0.1 mM salicylate and 700 ng/ml tetracycline, respectively. After incubation for 20h at 30 °C and 250 rpm, the cells were harvested, resuspended in lysis buffer to the same cell density, and lysed by sonication. HIV protease was detected by Western blot as in (A). (D) Production using over-expression conditions. Proteins were produced as described in chapter 7.7. E. coli KA13 cells producing the different capsids plus the tagged HIV protease-R10 were resuspended in lysis buffer to the same cell density, and lysed by sonication. In this case, prior to analysis, the samples containing AaLS-neg and AaLS-13 were diluted 5-fold relative to the sample containing AaLS-wt and that lacking any capsid variant to prevent overloading of the gels. HIV protease was detected by Western blot as in (A). The evolved AaLS-13 capsid substantially increases the yield of total soluble HIV protease-R10 under both sets of experimental conditions compared to AaLS-neg capsid, which is in turn superior to AaLS-wt and cells lacking any capsid variant.
The enhanced complexation capability of AaLS-13 confers a remarkable advantage to hosts producing the evolved protein, permitting cell growth at much higher inducer concentrations (Fig. 2.12), and hence much higher intracellular HIV protease concentrations (Fig. 2.11C and D), than is possible with AaLS-neg. Whereas cells producing AaLS-neg die when the tetracycline concentration exceeds 600 ng/ml, cells with AaLS-13 still permit growth at 1400 ng/ml (Fig. 2.12D).

![Graphs representing growth curves of cells transformed with plasmids encoding HIV protease-R10 and either AaLS-wt (green), AaLS-neg (blue), or the evolved AaLS-13 (red). Cultures were grown in liquid LB media at 30°C. Production of AaLS variants was induced with 100 µM salicylate. Expression of the protease gene was controlled by the tightly regulatable tetracycline promoter. In the absence of tetracycline (A) the protease is not produced, while addition of 450 ng/ml (B) or 700 ng/ml (C) tetracycline leads to increasing intracellular concentrations of the toxic enzyme. (D) Influence of HIV protease-R10 production on the growth of AaLS-13 containing cells. Production of AaLS-13 was induced with 100 µM salicylate. Production of HIV protease-R10 was induced with either 400, 600, 800, 1000, 1200, 1400 or 3200 ng/ml tetracycline. The growth curve obtained in the absence of tetracycline is shown in green.]
This growth advantage could conceivably result from a protective stress response induced upon overproduction of the evolved capsid. To examine the possibility of a protective stress response, we used real-time PCR to monitor the transcript levels of genes encoding several chaperones that are typically upregulated when cells are stressed in different ways. The transcript levels of groEL and dnaK, two typical stress-related genes, actually decrease somewhat for HIV protease-producing cells when the AaLS-13 is present. The ibpA gene encodes a molecular chaperone that affects protein solubility and inclusion body formation and is among the most upregulated genes in cells overproducing recombinant proteins (96). Although cells producing the AaLS-13 capsid plus protease have ca. 1.4-fold higher ibpA transcript levels than cells producing only the protease (Fig. 2.13D), this increase is considerably lower than that resulting from heat shock (Fig. 2.13A) or observed upon recombinant production of a variety of other proteins (typically >20-fold) (96). These small effects argue against a protective stress response induced by the evolved capsid as the origin of the growth advantage seen in Fig. 2.12C. The significantly higher overall yield of HIV protease-R_{10} (Fig. 2.11C and D) instead supports a mechanism involving direct association of the toxic protein and the evolved capsid.
**Figure 2.13.** Stress response induced by heat shock (A) or production of HIV protease-R_{10} in the presence and absence of AaLS-13 (B-D). Relative expression ratios of the stress response genes *groEL* (orange), *dnaK* (blue) and *ibpA* (green) were determined by real time PCR. Relative expression ratios were calculated from the efficiencies of the PCR reactions and the crossing point (CP) deviations between samples before and after induction or heat shock. The values were normalized using the gene *rpoD* as an endogenous control.

The higher loading capacity of AaLS-13 can be attributed to some or all of the seven point mutations that occurred during its optimization. The most striking feature of these alterations is that they substantially increase the net negative charge of each monomer. Although an aspartate at position 28 was mutated to a neutral glycine in the course of evolution, this change is more than offset by the addition of three new anionic residues (one glutamate and two aspartates) and the loss of three cationic residues (one lysine and two arginines). Significantly, none of these substitutions map to the capsid exterior (Fig. 2.14). Instead, two mutations, R127C and K131E, affect charged residues that project into the lumen of the capsid, and the remaining changes – D28G, R52C, T112S, V115D and A118D – alter residues located at subunit interfaces. Directed evolution thus appears to have improved on the original electrostatically-based encapsulation strategy by further optimizing the
luminal surface of the container for binding the positively charged cargo. In addition, some of the mutations at the subunit interfaces might also contribute by altering capsid dynamics.

Figure 2.14. Exterior and interior surfaces of a pentameric subunit of AaLS-wt (84). AaLS-neg mutations (R83E, T86E, T120E, Q123E) are shown in red, while new mutations in AaLS-13 (D28G, R52C, T112S, R127C, K131E) are in black. V115D and A118D are not visible from the perspective shown.

Capsids formed by AaLS-13 are very similar to AaLS-neg capsids. However, consistent with its enhanced loading capacity, the evolved protein forms somewhat larger spherical capsids, perhaps as a consequence of slight rearrangements at the intersubunit interfaces. Negatively stained electron microscopy images show that empty AaLS-13 capsids have an average diameter of 35.4 ± 3.2 nm, as compared to 28.6 ± 2.6 nm for AaLS-neg (Fig. 2.15). In the presence of the tagged protease, the diameter of AaLS-13 contracts slightly to 34.7 ± 3.4 nm. These values fall in the range previously seen for wild type and mutant lumazine syntheses (86), and are consistent with icosahedral capsids having T = 3 or T = 4 symmetry (i.e. 180 or 240 identical subunits). Icosahedral lumazine syntheses can adopt multiple assembly forms depending on the choice of buffer and pH (86). In addition to T=1 and T=3 capsids with diameters around 16 and 29 nm, respectively, a larger structure consistent with either an expanded T=3 or a T=4 state has been observed. Under the conditions used here, AaLS-neg assembles predominantly as conventional T=3 capsids, whereas the evolved variant appears to favor the larger expanded form,
both in the presence and absence of HIV protease, that. Although the observed shift in population can be ascribed to the increase in net negative charge on each monomer, detailed structural experiments will be needed to resolve whether it reflects a subtle subunit rearrangement or an increase in T number.

![Figure 2.15](image)

**Figure 2.15.** Negative stain electron microscopy pictures of AaLS-neg and AaLS-13, and AaLS-13 that had been coproduced with HIV protease-R10. Light regions correspond to protein. Scale bar, 100 nm.

Several lines of evidence in addition to the absence of mutations on the exterior of the capsid indicate that the tagged HIV protease is encapsulated by AaLS-13. For instance, no complex forms when AaLS-13 capsids and HIV protease-R10 are produced separately, mixed in vitro, and purified under standard conditions (Fig. 2.16). This result demonstrates that the association between these two proteins must take place before capsid assembly is completed. In addition, 1M NaCl fails to disrupt the complex, which rules out simple electrostatic binding of the R10-tagged protease to the capsid exterior (Fig. 2.16).
Figure 2.16. Association of HIV protease-R10 with AaLS-13 upon mixing in vitro compared to coproduction in vivo. SDS-PAGE (top) shows relative capsid concentrations. Western blot (bottom) shows relative amounts of HIV protease-R10 in the same samples, detected with a mouse anti-HIV protease antibody. Cleared cell lysate from a culture that coproduced AaLS-13 and HIV-protease-R_{10} is shown in lane one. Lane two sample consists of lysates derived from cultures, in which AaLS-13 and HIV-protease-R_{10} were each produced separately, and then mixed together in vitro. The lower band in the Western blot (bottom) most likely corresponds to untagged HIV protease and may result from cleavage by cellular proteases; it is proportionally much weaker in the coproduced sample, suggesting a protective effect of the capsid. Standard molecular weight marker is in the third lane. The fourth and fifth lanes contain samples in which the lysate from lane one was further purified under normal and high salt conditions, respectively. The similar band intensities in lanes four and five of the Western blot demonstrate that the complex remains intact even at high ionic strength, suggesting that the protease is caged within the capsid. The two rightmost lanes show samples in which the mixed lysate (from lane 2) was subsequently purified using Ni^{2+}-NTA affinity chromatography under normal and high salt conditions. The absence of detectable bands in the two rightmost lanes of the Western blot indicates that no complex forms when AaLS-13 and HIV protease-R_{10} are produced separately and mixed in vitro, suggesting that, during coproduction, complex formation takes place before capsid assembly is complete.

The most striking support for the localization of the protease within the capsid comes from cryo-electron microscopy (cryo-EM). Images of AaLS-13 obtained by cryo-EM (Fig. 2.17) reveal that many of the capsids formed upon coproduction with the protease possess darker interiors, and thus higher luminal protein contents, than capsids produced in the absence of the guest. This difference was confirmed by aligning and averaging hundreds of individual particle images from the cryo-electron micrographs (Fig. 2.17B) and quantitating the signal along cross-sections of the averaged images (Fig. 2.17D). In contrast to the protease-free sample,
however, the interiors of particles in the protease-containing sample display a broad distribution in cargo loading, ranging from empty to heavily loaded (Fig. 2.17A). To heighten the contrast with “empty” capsids (Fig. 2.17B, left), the images of AaLS-13 coproduced with the protease were therefore sorted into two populations based on alignment with reference capsids possessing representative “light” and “dark” interiors, which correspond to the low and high ends of the loading distribution, respectively. Although this crude sorting procedure ignores more subtle variation in loading, it improves the contrast significantly. 41% of the particles more closely matched the “light” reference, and the averaged image appears nearly empty (Fig. 2.17C, left, and 2.17D). The remaining 59% more closely matched the “dark” reference particle and the interior of the averaged image appears substantially denser (Fig. 2.17C, right, and 2.17D), strongly indicative of guest encapsulation. The apparent heterogeneity in capsid loading presumably depends on the relative rates of guest capture versus capsid assembly, and thus on the timing and extent of protease production.

**Figure 2.17.** Cryo-electron microscopy. (A) Cryo-electron micrograph of AaLS-13 coproduced with HIV protease-R10. Dark regions indicate the presence of protein. Scale bar, 100 nm. Capsids with representative “light” and “dark” interiors are indicated with white and black arrows, respectively. (B) Aligned and averaged cryo-electron micrographs of AaLS-13 produced alone (average of 636 particles) (left) and coproduced with HIV protease-R10 (average of 854 particles) (right). (C) Averaged images of AaLS-13 coproduced with HIV protease-R10 after sorting into two populations using reference capsids with representative “light” and “dark” interiors. Of the 854 capsids, 41% more closely match the “light” (left) and 59% the “dark” (right) reference. (D) Plots of averaged density across the cryo-EM images from (B) and (C) show that AaLS-13 capsids coproduced with protease (blue) have darker interiors than empty containers from samples lacking protease (black). After sorting, “dark” capsids in the protease-containing sample (red) have interiors with significantly higher signal intensity than either the “light” capsids from the same sample (green) or the sample lacking protease (black), consistent with localization of the protease inside the capsid.
2.3 Discussion

The encapsulation of R\textsubscript{10}-tagged HIV protease by AaLS-neg demonstrates that AaLS-neg capsids can serve as a container for guests other than GFP-R\textsubscript{10}, suggesting that this rationally designed system can be readily extended to any appropriately tagged guest. Sequestration of the toxic protease enabled optimization of the encapsulation system by directed evolution, yielding a capsid variant with an almost ten fold higher encapsulation efficiency than the starting AaLS-neg.

In its properties, AaLS-13 resembles other natural polyhedral protein containers, including wild-type lumazine synthase (23) and various other encapsulins (21) that form inclusion complexes with specific guest proteins. By encasing HIV protease within a thin protein shell, the engineered capsid protects the cell by placing a physical barrier between the enzyme and potential cytoplasmic substrates. This mode of inhibition is reminiscent of that used by the a-macroglobulins (97), which are found in the blood plasma of mammals and which regulate the activities of circulating proteases. Unlike a-macroglobulins, AaLS remains inside the cell and thus serves as an artificial prokaryotic organelle. Caging the protease provides a practical means of increasing its yield (Fig. 2.11C and D), suggesting a potentially general strategy for producing toxic enzymes.

Since assembly of the complex is achieved via a simple tagging strategy based on complementary Coulombic interactions between host and guest, extension of this approach to other difficult-to-produce proteins (or other biomolecules) should be straightforward. Given that some small molecules can pass through the pores in the shell (like the substrates of wild-type lumazine synthase (23)), encapsulation does not necessarily entail loss of biochemical function, so this system might also be employed to protect metabolically useful but unstable catalysts. Conceivably, more sophisticated tags that rely on specific binding sites on the capsid interior, like those utilized by natural encapsulins (21) to bind enzymes involved in the oxidative stress response, could be developed as well. The availability of multiple tags could be useful for isolating different cargo proteins in separate protein compartments or for co-localizing them within the same enclosure.
Molecular containers that encapsulate specific cargo have found many natural and nonnatural applications (33, 36, 37). Here we have built on this precedent, demonstrating the feasibility of regulating enzyme activity within the cell through selective complexation with an engineered capsid-forming protein. The ease with which this artificial host-guest system could be optimized by molecular evolution has implications for understanding how such symbiotic systems arise and develop in nature. In practical terms, it also paves the way for creating more sophisticated multifunctional protein assemblies that might extend the functional range of natural protein nanocompartments.
3 PROPERTIES OF AN EVOLVED PROTEIN CONTAINER

3.1 Introduction

Protein capsids that enclose specific cargo are used in nature for many specialized biological functions. Proteinaceous microcompartments act as simple organelles by sequestering specific metabolic processes (e.g. bacterial microcompartments like the carboxysome), protein folding is assisted in the cavity of chaperones, ferritin and viruses store metals or nucleic acids inside their capsules.

Protein capsids can also be used outside the cell. There is considerable interest in engineering molecular protein containers for diverse bio- and nanotechnological applications, where they serve as reaction vessels, nanotemplates and synthetic platforms for material synthesis, biosensors, and delivery vehicles.

Encapsulation of guests into protein nanocompartments often relies on mechanisms and recognition elements used in nature. The engineered encapsulation system developed in our laboratory and based on protein capsids of *Aquifex aeolicus* lumazine synthase (AaLS) shows that simple charge complementarity can also be used to drive encapsulation. Introduction of negatively charged glutamates on the inner surface of AaLS capsids (AaLS-neg) enables spontaneous encapsulation of cargo proteins such as GFP and HIV protease containing a complementary positively charged deca-arginine tag (R₁₀ tag). As shown in the previous chapter, it is possible to optimize this charge-driven encapsulation system by directed evolution. Several rounds of mutagenesis and selection yielded a capsid variant (AaLS-13) that could load higher amounts of HIV protease-R₁₀ than the starting capsid and thus permitted efficient cell growth even at high intracellular concentrations of the toxic enzyme.

Protection of cells against the toxic effects of the HIV protease by sequestration in AaLS-13 provided a practical means of increasing protease yields, suggesting a potentially general strategy for producing toxic enzymes. Although this application is biotechnologically relevant, it represents only one of many possible uses of this system. Encapsulation of unstable proteins that are prone to aggregation could
improve their production. But such a system could also be extended to sequester catalysts, and thereby gain more control over the process, or to serve as biocompatible delivery vehicles for biologically or medicinally relevant molecules.

To realize these opportunities, detailed characterization of this encapsulation system is required. Here, the properties of AaLS-13, the evolved container, are compared with those of AaLS-neg. Specifically, the encapsulation of other guests, the specificity of this process in vivo, and structural properties of the complexes, including stability of the host-guest assemblies, are investigated. Our results show that virtually any positively charged protein can be encapsulated by AaLS-13 in principle. Nevertheless, the evolved capsid exhibits some selectivity towards appropriately tagged guests like HIV protease-R10. Moreover, although AaLS-13 assembles into multiple quaternary structures, binding of the guest appears to stabilize the capsid state.

3.2 Results

3.2.1 In vivo guest encapsulation

Cytosolic proteins and HIV protease-R10

Chapter 2 shows that the tagged HIV-protease-R10 is efficiently encapsulated by AaLS-13 when the two proteins are coproduced in E. coli cells. Since the encapsulation of guests into AaLS-13 capsids is mainly based on electrostatic interactions, any highly positively charged macromolecule in the cell could in principle become encapsulated by the evolved capsids.

To determine the in vivo selectivity of our system we used mass spectrometry. Capsids AaLS-wt, AaLS-13, and AaLS-13 that had been coproduced with HIV protease-R10 were purified by Ni2+-affinity, anion exchange and size exclusion chromatography. The purified capsids were digested with trypsin, and then analyzed by LC-MS/MS. Proteins detected in the samples were quantified based on the label-free Top 3 Protein Quantification (T3PQ) method (98) (Fig. 3.1). Whereas wild-type lumazine synthase associates predominantly with the highly abundant
50S ribosomal protein L2 and proteins like β-lactamase and nahR that are overproduced from the same plasmid as the capsid, the evolved AaLS-13 variant co-purifies as expected with several positively-charged ribosomal proteins, including L2, S4, and a few others that have pI values >10 and are among the most abundant cytosolic proteins in *E. coli* (99).

Nevertheless, appropriately tagged guests molecules can effectively compete with these abundant cytosolic proteins. When AaLS-13 is co-produced with the R₁₀-tagged HIV protease, the capsids additionally contain substantial amounts of HIV protease. For example, when protease production was induced with 5000 ng/ml tetracycline, the amount of encapsulated HIV protease is more than twice that of the most abundant ribosomal protein. The capsids appear to contain somewhat higher amounts of the ribosomal proteins than AaLS-13 produced alone, which probably reflects a stabilizing effect of the protease on the capsid structure (see below), leading to less opening and closing of the capsid and consequently to less loss of encapsulated cargo during isolation and production as compared to the empty capsids. Given the comparatively low concentration of HIV protease, these results show that the engineered system exhibits selectivity towards the intended guest.
Figure 3.1. Mass spectrometric analysis of AaLS capsids. The bar graph shows the relative amounts of the most abundant proteins that copurify with the capsids AaLS-wt (yellow), AaLS-13 (purple), and AaLS-13 coproduced with HIV protease-R_{10} (red). Abbreviations (pI value as calculated from the amino acid sequence by ProtParam (100)): capsid, AaLS variant; HIV-R10, HIV protease-R_{10} (pI = 11.1); RplB, 50S ribosomal protein L2 (10.9); RpsD, 30S ribosomal protein S4 (10.1); RpsR, 30S ribosomal protein S18 (10.6); RplO, 50S ribosomal protein L15 (11.2); RpsK, 30S ribosomal protein S11 (11.3); RplE, 50S ribosomal protein L5 (9.5); Bla, β-lactamase TEM (5.5); RplQ, 50S ribosomal protein L17 (11.1); RpsE, 30S ribosomal protein S5 (10.1); RpsL, 30S ribosomal protein S9 (10.9); RpsC, 30S ribosomal protein S3 (10.3); GatZ, D-tagatose-1,6-bisphosphate aldolase subunit gatZ (5.5); DeaD, cold-shock DEAD box protein A (8.8); NahR, HTH-type transcriptional activator NahR (6.6); RpsM, 30S ribosomal protein S13 (10.8); RpsL, 30S ribosomal protein S12 (10.9); RplN, 50S ribosomal protein L14 (10.4).

GFP variants

Although the MS analysis shows that diverse highly positively charged macromolecules can be encapsulated by AaLS-13 capsids, it is not clear how the distribution of positive charges on the guest influences encapsulation efficiency. To examine these issues, encapsulation of four GFP variants possessing different net positive charge and charge distributions were studied with both AaLS-neg and AaLS-13 (Fig. 3.2).

Wild type GFP, which has a calculated pI of 5.7, is the variant with the fewest positive charges. When the R_{10} tag is appended to its C-terminus the pI increases to
The third guest molecule is a highly positively charged GFP variant in which 29 surface exposed residues were mutated to positively charged amino acids (10I), yielding a protein with a theoretical net charge of +36 and a calculated pI of 10.4. The R$_{10}$ tag was also attached to this superpositive GFP (GFPpos) resulting in a variant with a pI of 10.9.

The GFP variants were each coproduced with AaLS-neg and AaLS-13, and the capsids were isolated by Ni$^{2+}$-affinity chromatography. The encapsulation efficiencies were determined by measuring GFP fluorescence associated with the purified capsids. As shown previously (8I), the R$_{10}$ tag is very important for encapsulation of normal GFP into AaLS-neg. Only small amounts of wild type GFP copurify with AaLS-neg, but appending the positively charged R$_{10}$ tag to the guest leads to a greater than ten-fold increase in capsid-associated fluorescence (Fig. 3.3). Surprisingly, a single R$_{10}$ tag is less effective for encapsulation by the evolved AaLS-13 variant. Wild type GFP is encapsulated by AaLS-13 slightly more efficiently than by AaLS-neg, but attaching the R$_{10}$ tag results in less than four-fold higher binding to AaLS-13. By measuring the total amount of GFP that is present in the cleared cell

**Figure 3.2.** Encapsulation of differently charged GFP variants. The numbers indicate the estimated numbers of GFP molecules encapsulated per T=3 AaLS capsid.
lysates, the percentage GFP that is encapsulated was determined (Fig. 3.3C). The trend of these values correlates with the GFP amounts associated per capsid (Fig. 3.3A and B). However, even at best the highest fraction of captured GFP only accounts for ca. 0.4% of the total GFP produced. This was seen for the combination of AaLS-neg and GFP-R10.

GFPpos copurifies with AaLS-neg to a similar extent as GFP-R10, and addition of the R10 tag to this construct results in no further improvement (Fig. 3.4). The evolved capsid variant AaLS-13 is the better host for GFPpos; it binds about four times more GFPpos than does AaLS-neg. The amounts of GFPpos encapsulated by AaLS-13 are also significantly higher than the amounts of GFP-R10, indicating a preference for guests with a highly positively charged surface. Attachment of the R10 tag to the superpositive GFP does not lead to further improvement in encapsulation efficiency by AaLS-13; the number and distribution of positive charges on GFPpos itself seems to be sufficient for effective loading of the evolved capsids. The percentage of total AaLS-associated GFP (Fig. 3.4C) agrees with the GFP amounts bound per capsid (Fig. 3.4A and B). It is generally higher than with GFP-R10 as guest, and AaLS-13 encapsulates more than 5% of total GFPpos present in the cell.

Previous measurements determined that AaLS-neg encapsulates three to four GFP-R10 molecules per T=3 capsid (3.8 ± 0.8, (B1)). Assuming that AaLS-neg binds a similar amount of GFP-R10 in the experiments described here, the number of encapsulated guests can be estimated for the different combinations examined (Fig. 3.2). GFPpos exhibits a two fold lower fluorescence per mol protein, which was also taken into account. Based on these considerations AaLS-neg binds similar numbers of supercharged GFP and GFP-R10 (three to four GFPpos per capsid), and the evolved variant AaLS-13 encapsulates one to two molecules of GFP-R10. However, as for HIV protease-R10, AaLS-13 exhibits a superior encapsulation capability for GFPpos. It shows four-fold higher loading than AaLS-neg encapsulating about 13 GFPpos molecules per T=3 capsid. In comparison, each AaLS-13 capsid binds approximately seven R10-tagged HIV protease dimers, which represents a five- to ten-fold improvement over AaLS-neg.

For efficient encapsulation into AaLS-neg and AaLS-13 the presence of a sufficient number of positive charges is crucial. In the case of a non-positively
charged guest like GFP, attachment of the R$_{10}$ tag is crucial for binding to the AaLS variants. However, the presence of the R$_{10}$ tag results in no further improvement if the guest is already highly positively charged like the GFP variant GFPpos. The evolved capsid AaLS-13 seems to prefer guests with an even higher number of positive charges. In contrast to AaLS-neg, it shows an increase in encapsulation efficiency from the single-tagged GFP to the GFPpos with a highly positively charged surface. Given that the HIV protease, against which AaLS-13 was evolved, is already positively charged without the R$_{10}$ tag (calculated pI of 7.8) and forms a dimer with two R$_{10}$ tags, this result is not unexpected. But most important, the evolutionary optimization of AaLS-13 for higher encapsulation efficiency of HIV protease-R$_{10}$ translates to other highly positively charged proteins, such as GFPpos. As a consequence, efficient encapsulation of new positively charged guests should be straightforward.

![Figure 3.3](image)

**Figure 3.3.** Encapsulation efficiencies of GFP and GFP-R$_{10}$ by AaLS-neg and AaLS-13. (A) The GFP specific fluorescence associated with purified AaLS capsids from different purifications (red, orange and yellow) was measured at a total protein concentration of 0.05 mg/ml. (B) Fluorescence values from (A) were normalized to samples AaLS-neg+GFP-R$_{10}$ and averaged. (C) GFP-specific fluorescence associated with purified AaLS capsids expressed as percentage of the total GFP-specific fluorescence in cell lysates. The average of the different purifications (red, orange and yellow) is shown in blue.
Figure 3.4. Encapsulation efficiencies of GFPpos and GFPpos-R<sub>10</sub> by AaLS-neg and AaLS-13. (A) The GFP specific fluorescence associated with purified AaLS capsids from different purifications (blue, purple) was measured at a total protein concentration of 0.05 mg/ml and corrected for the intrinsic 1.75 times lower fluorescence of GFPpos compared to GFP. (B) Fluorescence values from (A) were normalized to samples AaLS-neg+GFPpos-R<sub>10</sub> and averaged. (C) GFP-specific fluorescence associated with purified AaLS capsids expressed as percentage of the total GFP-specific fluorescence in cell lysates. The average of the different purifications (blue, purple) is shown in red.

Preliminary work in the Woycechowsky laboratory shows that AaLS-13 system can be extended to other guests and encapsulate an active enzyme (K. Woycechowsky, personal communication). AaLS-13 was coproduced with a monomeric 55 kDa esterase from *Geobacillus stearothermophilus* (<sup>102</sup>) bearing the R<sub>10</sub> tag. The esterase copurified with the capsids in Ni<sup>2+</sup>-NTA and gel filtration chromatography. Importantly, the encapsulated enzyme appears to retain its hydrolytic activity. Thus, it shows seven-fold higher specific activity for the hydrolysis of paranitrophenylacetate. In contrast, with purified, HIV protease-loaded AaLS-13 only negligible protease activity was observed, presumably because the smallest substrate that is efficiently cleaved by HIV protease, a hexa-peptide, is too large to diffuse through the capsid pores.

The fact that an enzyme can be incorporated into the capsid and retain its activity suggests a variety of applications. For example, an encapsulated enzyme could catalyze the synthesis of nanomaterials, such as a polymer inside the cavity. This would yield homogenous particles defined by the size of the protein capsid, something that is difficult to accomplish without a template. The pores of AaLS capsids could also be used to control access of substrates to the caged enzyme, thus...
achieving selectivity. Further, multienzyme complexes could be engineered by compartmentalizing different catalysts.

3.2.2 Stability of the capsids and host-guest complexes

The evolved lumazine synthase variant AaLS-13 contains seven additional mutations compared to AaLS-neg. How do the accumulated changes alter capsid structure? In Chapter 2 it was shown that AaLS-13 forms slightly larger capsids than AaLS-neg. Here we investigate the effects of the mutations on a variety of other properties, including stability, secondary structure, capsid assembly, and dissociation of various host-guest complexes.

Capsid stability

Capsid assembly was investigated by size exclusion chromatography and showed that the mutations accumulated during directed evolution influence capsid assembly of AaLS-13. AaLS-neg elutes as a capsid (9 ml), whereas AaLS-13 elutes in two fractions, the first corresponding to intact capsids (9 ml) and the second to pentamers (15 ml) (Fig. 3.5A). Such incomplete assembly is not unprecedented; lumazine synthases from other organisms, such as S. cerevisiae, form pentamers exclusively (82). Isolation of the AaLS-13 capsid fraction and reinjection on the size exclusion column after three days gave a similar elution profile when AaLS-13 was produced without a tagged guest: capsids had partially disassembled to pentamers (Fig. 3.5B). Isolated pentamers did not assemble to capsids (Fig. 3.5D and E). However, when AaLS-13 capsids that are loaded with HIV protease-R₁₀ are isolated and reinjected, the capsids remained stable and did not disassemble (Fig. 3.5C). These observations indicate that the AaLS-13 capsid is more labile than its progenitor AaLS-neg and that guest loading with HIV protease-R₁₀ stabilizes the AaLS-13 capsid.
Figure 3.5. Quaternary structure of AaLS variants. (A) Size exclusion chromatography traces of Ni\textsuperscript{2+}-affinity purified AaLS-neg (red), AaLS-13 (green), AaLS-13+HIV proteas-R\textsubscript{10} (black), and pentameric ScLS (blue) as a marker run over a sephacryl S-300 column. (B) Chromatogram of the isolated AaLS-13 and (C) AaLS-13+HIV proteas-R\textsubscript{10} capsid fractions reinjection on the sephacryl S-300 column after three days. (D) Chromatogram of the isolated AaLS-13 and (E) AaLS-13+HIV proteas-R\textsubscript{10} pentamer fractions reinjection on the sephacryl S-300 column after three days.

The structural rearrangements that give rise to the observed quaternary structure heterogeneity are rather subtle. CD spectroscopy revealed no changes in secondary structure in AaLS-13 capsids compared to AaLS-neg (Fig. 3.6A). Binding of HIV protease-R\textsubscript{10} also had no detectable effect on secondary structure of AaLS-13 capsids. However, AaLS-13 pentamers show a slightly different spectrum with a more pronounced minimum at 210 nm, indicating slightly different subunit conformations when assembled as isolated pentamers and as capsids. The thermal stability of the evolved AaLS variant decreased somewhat from $T_M = 97$ °C for AaLS-neg to $T_M = 78$ °C for AaLS-13 capsids and guest binding does not seem to have a significant effect (Fig. 3.6A-D).
Figure 3.6. CD spectroscopy. (A) Far-UV CD spectra of AaLS-13 capsids (red) and pentamers (yellow), AaLS-13+HIV protease-R
10 capsids (green) and pentamers (black), and AaLS-neg capsids (purple). The spectra were recorded in 50 mM sodium phosphate buffer (pH 8) containing 200 mM NaCl at a protein concentration of 10 µM. (B) Temperature-dependent unfolding and (C) refolding was monitored using the ellipticity at 222 nm. (D) Melting temperatures were calculated from the denaturation curves by determining the point of reflection using the first derivative.

Nevertheless, capsid formation is crucial for binding of the HIV protease-R
10 to AaLS-13. ELISA of the individual size exclusion fractions showed that the HIV protease elutes solely with the peak corresponding to the capsid. Size exclusion chromatography followed by ELISA detection also demonstrated that binding of the HIV protease-R
10 to both AaLS-neg and AaLS-13 capsids is very tight as no release of the protease is observed even after 20 days (Fig. 3.7).
**Figure 3.7.** Dissociation of HIV protease-R<sub>10</sub> from AaLS-neg and AaLS-13 capsids. (A) Aliquots of purified AaLS-neg+HIV protease-R<sub>10</sub> and (B) AaLS-13+HIV protease-R<sub>10</sub> were injected on a superose 6 size exclusion column after different time periods as indicated in the legend. (C) and (D) ELISA detection of HIV protease in the individual fractions from the size exclusion runs of (A) AaLS-neg+HIV protease-R<sub>10</sub> and (B) AaLS-13+HIV protease-R<sub>10</sub>, respectively.

**Stability of GFP complexes**

To compare the leakiness of the AaLS capsids with different guests, dissociation of the GFP-R<sub>10</sub>, GFPpos and GFPpos-R<sub>10</sub> variants from AaLS-13 and AaLS-neg capsids over time was studied by size exclusion chromatography. Capsids were coproduced with the GFP variants and then isolated by Ni<sup>2+</sup>-affinity chromatography. Aliquots of Ni<sup>2+</sup>-affinity purified samples were subjected to size exclusion chromatography after various time periods, and GFP in the individual fractions was detected by fluorescence.

AaLS-neg coproduced with GFP-R<sub>10</sub> elutes in at least three quaternary states from the size exclusion column at volumes corresponding to T=3 capsids (180mers, 9 ml), T=1 capsids (60mer, 11 ml) and a quaternary state that lies between the size expected for a pentamer and a 60mer, which is presumably a partially assembled capsid fragment (14.5 ml) (Fig. 3.8A). In these experiments GFP-R<sub>10</sub> initially
associates almost exclusively with this capsid fragment but dissociates with a half-life of about 20 h (free GFP-R10 elutes at 21.5 ml) (Fig. 3.8C and E). Upon dissociation of GFP-R10 the capsid fragment seems to assemble into T=1 capsids. In contrast, when HIV protease-R10 is used as the guest, AaLS-neg forms fully assembled capsids. The protease remains tightly associated with these capsids even over extended time periods (Fig. 3.7).

GFP-R10 does not disturb capsid assembly of AaLS-13, which elutes as a single peak corresponding to T=3 capsids, and remains bound to these capsids for long periods (Fig. 3.8B and D). For example, about 65% of the guest remains associated with AaLS-13 even after 150 h, whereas GFP-R10 almost completely dissociates from AaLS-neg after 50 h (Fig. 3.8E and F). The much slower dissociation is presumably due to actual encapsulation of GFP-R10 by AaLS-13, as opposed to binding of the guest to only partially assembled AaLS-neg capsids.
Figure 3.8. Dissociation of GFP-R\textsubscript{10} from AaLS-neg and AaLS-13 capsids. (A) Aliquots of purified AaLS-neg+GFP-R\textsubscript{10} and (B) AaLS-13+GFP-R\textsubscript{10} were injected on a superose 6 size exclusion column after different time periods as indicated in the legend. (C) and (D) GFP fluorescence detection of GFP-R\textsubscript{10} in the individual fractions from the size exclusion runs of (A) AaLS-neg+GFP-R\textsubscript{10} and (B) AaLS-13+GFP-R\textsubscript{10}, respectively. (E) and (F) The total GFP-R\textsubscript{10} fluorescence of the run was set to 1 and relative fluorescence of AaLS-associated GFP-R\textsubscript{10} (blue) and free GFP-R\textsubscript{10} (red) were determined. Results for sample AaLS-neg+GFP-R\textsubscript{10} are shown in (E). Experiments with AaLS-13+GFP-R\textsubscript{10} exhibited a greater variation, therefore data from four purifications are shown and were exponentially fit using equation $m_1 + m_2 \times \exp(-m_3 \times x)$.

Upon coproduction with the superpositively charged GFP variant, GFPpos, AaLS-neg initially shows the same quaternary states as with GFP-R\textsubscript{10} (Fig. 3.9A). However, GFPpos does not dissociate from AaLS-neg capsid segments (Fig. 3.9C). Instead, it stays bound, and the capsid fragments ultimately assemble to T=1 capsids. This process is almost complete after 48 h and GFPpos remains bound to the capsids even after 164 h.

The evolved capsid AaLS-13 elutes as expected for a T=3 capsid and the guest GFPpos coelutes exclusively with the AaLS-13 capsids (Fig. 3.9B and D). No release of GFPpos is observed even after 164 h, as seen with HIV protease-R\textsubscript{10}. The R\textsubscript{10}-tagged GFPpos behaves like untagged GFPpos.
Highly positively charged proteins like the positively charged GFP variant and the HIV protease-R₁₀ bearing two R₁₀ tags bind very tightly to both AaLS-neg and AaLS-13 capsids. However, attachment of only one R₁₀ tag to GFP results in relatively fast guest release and impairment of capsid assembly in the case of AaLS-neg. The evolved variant is able to retain GFP-R₁₀ substantially longer.

Figure 3.9. Dissociation of GFPpos from AaLS-neg and AaLS-13 capsids. (A) Aliquots of purified AaLS-neg+GFPpos and (B) AaLS-13+GFPpos were injected on a superose 6 size exclusion column after different time periods as indicated in the legend. (C) and (D) GFP fluorescence detection of GFPpos in the individual fractions from the size exclusion runs of (A) AaLS-neg+GFPpos and (B) AaLS-13+GFPpos, respectively.
3.3 Discussion

The loading of the evolved AaLS-13 capsids is based on charge complementarity between host and guest. This principle appears to be very general, allowing straightforward loading of the capsids with diverse positively charged proteins. Efficient and stable encapsulation of GFP variants and an active esterase demonstrates that this simple system, which was optimized for HIV protease, can be readily extended to other guests. These results thus pave the way to new applications. For example, confinement in AaLS capsids could be used to stabilize fragile proteins and enzymes by improving folding or sequestration away from proteases. Caging multiple enzymes simultaneously within the capsid might enable catalysis of cascade reactions. Due to substrate channeling, such processes might become more efficient.

Because loading is based on simple electrostatic complementarity between host and guest, any positively charged protein can be encapsulated, including abundant cytosolic proteins with high pI values. These molecules can compete with the tagged guest, with efficiencies of encapsulation depending on their binding affinities and concentrations during capsid assembly. Nevertheless, the preferential incorporation of tagged HIV protease suggests the system exhibits some selectivity – particularly given the comparatively low concentration of the toxic protein in the cell relative to highly abundant ribosomal proteins. One way to enhance the specificity of guest binding would be the development of more sophisticated tags that exhibit high affinity for sites on the capsid interior. Natural encapsulins use such a strategy to sequester enzymes that carry a short C-terminal peptide tag, which binds to a conserved site on the capsid interior (21). In the case of AaLS such a tag could be based on the recognition elements used by lumazine synthase from *Bacillus subtilis* for natural encapsulation of the enzyme riboflavin synthase (23), once more structural information of the complex becomes available.

Another possibility to achieve specific encapsulation of only the intended guests would be the controlled loading in vitro. Similar to most viruses and other capsid-forming proteins lumazine synthases show multiple assembly states depending on the experimental conditions and upon mutations to the structure (86, 103). Our
results show that AaLS-13 capsids can disassemble into pentamers, especially in the absence of a guest. Typically, viruses can be loaded with non-natural guests by capsid disassembly and reassembly around the guest. Such processes are sometimes supported or induced by the guest itself (51-54, 104). By controlled disassembly of AaLS-13 and guest-induced reassembly, it might be possible to load these capsids with novel cargo in vitro. By carrying out these experiments outside the cell, competition with unwanted guests would be eliminated and a wide range of guests could be considered, including non-protein cargo. These possibilities are considered in the next chapter.
4 **IN VITRO ENCAPSULATION OF PROTEINS**

4.1 **Introduction**

The encapsulation of cargo molecules has found many natural and nonnatural applications in fields ranging from medicine to bio- and nanotechnology and material sciences. Engineered encapsulation systems often take advantage of the controlled in vitro loading so that virtually any desired guest can be chosen. Such systems typically use virus capsids (36, 37). Many viral capsids have the ability to form a range of different assemblies, i.e. they show polymorphism (103, 105, 106). Transitions between the different structures are dynamic and can be utilized in the laboratory for the controlled disassembly and reassembly around intended cargo by changing the pH, ionic strength, or metal ion concentrations. Reassembly can be induced by the guest (104), with the guest's properties sometimes even determining the resulting assembly state (49, 107).

The cowpea chlorotic mottle virus (CCMV) is among the best studied and most often used viral capsids. It shows reversible pH- and metal ion-dependent gating (108), providing an opportunity for controlled encapsulation and release of cargo. When the pH is raised above 6.5, in the absence of metal ions, the CCMV capsid swells by ca. 10% and 60 pores with a diameter of 2 nm open. These openings allow free exchange of large molecules between the capsid interior and the medium. Aqueous tungstate (WO$_4^{2-}$) and vanadate (VO$_{3+}$), and an organic polymer (polyanetholesulphonic acid) were loaded into CCMV in this way and then trapped when the pH was lowered to close the capsid pores (79). Lowering the pH simultaneously induced oligomerization of the tungstate and vanadate to form mineral nanoparticles of well-defined size within the viral capsid. CCMV capsids completely disassemble into dimers when the pH is further raised above 7.5. Complete capsid disassembly is reversible, which enables encapsulation of larger guests such as proteins (71, 76) or DNA-templated chromophore assemblies (109). In one example, individual horseradish peroxidase enzymes were sequestered into CCMV allowing examination of enzymatic activity at the single-molecule level (71).
Other viral cages and nonviral protein capsids like ferritin can also be disassembled and reassembled to encapsulate modified gold nanoparticles, quantum dots, or paramagnetic contrast agents for applications in bioimaging (51, 52, 54, 91), and proteins and drugs for delivery purposes (64).

Like CCMV, lumazine synthase from Aquifex aeolicus forms icosahedral protein capsids and exhibits polymorphism depending on conditions. Lumazine synthase capsids have been utilized for the size-constrained synthesis of metal oxide nanoparticles. The metal ions entered the capsid by diffusion through pores in the AaLS capsid shell. For encapsulation of larger guests into AaLS a different strategy has to be adopted. In our experiments of GFP and HIV protease, the engineered AaLS capsids were loaded by coproduction of the host with appropriately tagged versions of the guest in the cytosol of E. coli cells. Host-guest interaction is driven by engineered electrostatic interactions and encapsulation of the proteins presumably takes place during capsid synthesis and assembly inside the cell. It was possible to optimize this system by directed evolution (Chapter 2). The evolved variant AaLS-13 stably and efficiently encapsulates highly positively charged proteins and enzymes, bearing either a deca-arginine tag (R10 tag) or possessing a highly positively charged surface (Chapter 3), within the cells. The loading of AaLS capsids in vitro would be an appealing alternative.

Reminiscent of the viral capsids described above AaLS-13 shows quaternary structure heterogeneity. Especially when no guest is present its capsids can disassemble into pentamers (Chapter 3). Apparently, AaLS-13 capsids are labile in the absence of cargo, but become stabilized by guest binding. Here we show that these structural transitions can be exploited for controlled disassembly and reassembly to efficiently encapsulate a positively charged GFP variant (GFPpos) in vitro. Intact capsids with extremely high guest loadings can be obtained under favorable conditions. In vitro loading should allow controlled and specific encapsulation of a variety of protein and non-protein guests into AaLS-13, such as nanoparticles with useful optical or magnetic properties, medicinally relevant compounds or multiple enzymes.
4.2 Results

As shown in Chapter 3 capsids of the evolved variant AaLS-13 are more labile in the absence of guest and partly disassemble into pentamers. We also discovered that addition of EDTA to AaLS variants after Ni\textsuperscript{2+}-affinity purification and subsequent storage in an EDTA containing buffer is important for obtaining intact capsids. In experiments where EDTA was omitted the fraction of disassembled capsids is greatly increased. We wondered whether we could exploit these properties to achieve selective loading of AaLS-13 capsids with guests in vitro.

To test this idea, AaLS-13 capsids were produced without a guest and purified by Ni\textsuperscript{2+}-affinity chromatography. The sample was split into portions: 5 mM EDTA was added to one half, but omitted from the other. Both samples were subjected to size exclusion chromatography using buffer without EDTA. As expected they contained disassembled capsids, and the sample to which EDTA had been added after Ni\textsuperscript{2+}-affinity purification showed substantially more capsid, demonstrating the stabilizing effect of EDTA (Fig. 4.1A). In this sample the capsid fraction eluting at a volume of 8 ml from the sephacryl S-300 column constitutes the highest peak. Additional peaks eluting at 14 ml and 10 ml correspond to the size of a pentamer and a larger, partially assembled capsid fragment (between 15 to 50mer). But presumably a whole range of assembly states is present, ranging from monomers (should elute at 18.5 ml) to 180mers and exact identification of the various species is not possible from the size exclusion chromatograms. In the sample to which no EDTA was added after Ni\textsuperscript{2+}-affinity purification almost no capsids are present; instead smaller assemblies are formed with the major species eluting at 11 ml.

Fractions corresponding to all quaternary states smaller than capsids (9.5 ml-18.5 ml) were pooled. These disassembled AaLS-13 capsids were mixed with superpositive GFP in molar ratios of AaLS-13 monomer to GFP\textsuperscript{pos} of 5:1 and 1:1 in the presence of 5 mM EDTA to initiate reassembly around the positively charged guest.

The reassembly process was monitored by injecting aliquots of the sample on a superose 6 size exclusion column at different time intervals. A molar ratio of 1:1 led to almost complete precipitation of the proteins. A 5:1 ratio resulted in successful
reassembly of AaLS-13 capsids. After 44 h the sample to which EDTA had been added after the Ni$^{2+}$-affinity column was essentially fully reassembled to capsids (9 ml); no further changes were seen after another 119 h (Fig. 4.1B). In the sample without EDTA assembled capsids also constituted the largest fraction of the mixture, but smaller quaternary states were also observed (Fig. 4.1C). The latter eluted at volumes corresponding to capsid fragments (15 ml, corresponding to a MW of 150 to 300 kDa), pentamers (MW of 90 kDa, 18.5 ml). When no guest but only EDTA was added to disassembled AaLS-13, capsids also reassembled, but to a lesser extent (Fig. 4.1D). Incomplete assembly was especially pronounced for the sample to which no EDTA had been added after Ni$^{2+}$-affinity purification. Thus, guest binding is not essential for reassembly of AaLS-13 but it supports capsid formation, enabling more complete reassembly.

Fluorescence measurements of the individual fractions from the size exclusion runs described show that in the sample treated with EDTA nearly complete capsid reassembly is accompanied by almost exclusive binding of the GFPpos to these capsids (Fig. 4.1E). In the sample without EDTA GFPpos is associated mainly with two AaLS-13 assembly states, with intact capsids on the one hand, and with the partially assembled capsid fragments eluting at 15 ml (Fig. 4.1F) on the other. The latter resemble the complexes formed upon coproduction of AaLS-neg with GFP variants in vivo, which were described in Chapter 3 and elute at a similar volume from the size exclusion column. This experiment suggests that better in vitro loading can be achieved when EDTA is added to AaLS-13 after Ni$^{2+}$-affinity purification.
Figure 4.1. AaLS-13 disassembly, reassembly and loading. (A) Size exclusion chromatography traces of Ni$^{2+}$-affinity purified AaLS-13 treated with EDTA (blue) and without (red) run in buffer without EDTA on a sephacryl S-300 column. (B) and (C) Aliquots of the reassembly mixtures with a AaLS-13 to GFPpos ratio of 5:1 were injected on a superose 6 size exclusion column in buffer containing EDTA after different time periods. (B) Reassembly mixture of the sample (A, blue), which had been treated with EDTA, injected after 44 h (purple) and 163 h (blue). (C) Reassembly mixture of the sample (A, red), which had not been treated with EDTA, injected after 39 h (red) and 158 h (orange). (D) Reassembly of AaLS-13 without GFPpos of the sample treated with EDTA and injected after 181 h (blue), and the sample not treated with EDTA and injected after 179 h on the superose 6 column in buffer containing EDTA (red). (E) and (F) GFP fluorescence detection of GFPpos in the individual fractions from the size exclusion runs of (B) and (C), respectively.

The structural integrity of the reassembled AaLS-13 capsids was confirmed by negative staining electron microscopy (Fig. 4.2). Importantly, the guest protein is observed in these images, bound inside the lumen of the capsid. Images of capsids assembled with and without GFPpos showed that the reassembled capsids are similar in size and shape to the AaLS-13 capsids described in Chapter 2, Fig. 2.15, which had not gone through the procedure of disassembly and reassembly. Like the latter, the reassembled capsids have a diameter of ca. 35 nm, which is consistent with icosahedral capsids having $T = 3$ or $T = 4$ symmetry. Only capsids assembled in the presence of GFPpos contain large amounts of protein within the cavity, suggesting very high guest loading. The guest appears to localize along the capsid walls. In comparison, when AaLS-13 was loaded in vivo with HIV protease-R$_{10}$
(Chapter 2) the luminal protein content (seven HIV protease-R₁₀ dimers per T=3 capsid) was insufficiently high to be visualized in negative staining EM images of the complex. As a result cryo-EM had to be carried out to provide evidence for encapsulation of HIV protease-R₁₀.

![Image of electron microscopy images](image)

**Figure 4.2.** Negative staining electron microscopy images of AaLS-13 capsids reassembled with (left) and without GFPpos (right). Light regions correspond to protein. Scale bar, 100 nm.

The number of GFPpos molecules loaded per AaLS-13 capsid was estimated by fluorescence. Using a standard curve generated with free GFPpos of known concentrations we find that ca. 63 molecules of GFPpos are encapsulated per T=3 capsid in the sample treated with EDTA and ca. 40 in the sample without EDTA. These yields are about four fold higher than the loading obtained by coproduction of AaLS-13 and GFPpos in vivo (Chapter 3).

The amount of GFPpos encapsulated per AaLS-13 capsid is expected to depend on the ratio in which capsid and GFPpos are mixed for reassembly. To test this hypothesis, in vitro loading with different molar ratios of AaLS-13 to GFPpos was investigated. As before, Ni²⁺-affinity purified AaLS-13 was mixed with EDTA, subjected to size exclusion chromatography in buffer without EDTA, and fractions corresponding to disassembled capsids were pooled (10 ml – 18 ml) and mixed with GFPpos in molar ratios of 5:1, 15:1 and 45:1 in the presence of EDTA.

In contrast to the experiments described above only a small fraction of the disassembled AaLS-13 reassembled into capsids (Fig. 4.3). The best yields were obtained for the sample in which AaLS-13 and GFPpos were mixed in a ratio of 5:1 (Fig. 4.3A and D). These reassembled capsids contained GFPpos, but most of the
GFPpos was associated with the two smaller and more abundant quaternary states that eluted at volumes corresponding to capsid fragments and pentamers. Both structure and assembly seem to have been severely impaired in this experiment, perhaps because AaLS-13 was kept too long in buffer without EDTA following size exclusion chromatography.

Nevertheless, mixing GFPpos with size exclusion fractions corresponding to intact capsids (7.5 ml – 9 ml on the sephacryl S300 column) in a ratio of 5:1 (AaLS-13:GFPpos) and in the presence of EDTA resulted in efficient encapsulation of GFPpos (Fig. 4.4). Negative staining EM images of this sample show intact and highly filled AaLS-13 capsids and quantification revealed that about 30 GFPpos molecules are encapsulated per T=3 capsid. The yield is lower than in the first experiment, where disassembled capsid segments were used for loading, but it may reflect the variation of the experiment itself. Capsid assembly/disassembly is apparently a highly dynamic process in the absence of EDTA, allowing efficient encapsulation upon transient opening of the capsid. Once guest is bound, the entire assembly becomes substantially more stable as leakage is not observed from the host-guest complexes over 20 days.
Figure 4.3. AaLS-13 reassembly and loading with different AaLS-13 to GFPpos ratios. (A), (B) and (C) Aliquots of the reassembly mixtures with a AaLS-13 to GFPpos ratio of 5:1, 15:1 and 45:1 were injected on a superose 6 size exclusion column in buffer containing EDTA after different time periods. (A) Injection of the 5:1 reassembly mixture after 15 h (red), 36 h (blue) and 135 h (green), (B) the 15:1 mixture after 16 h (red), 37 h (blue) and 137 h (green) and (C) the 45:1 mixture after 18 h (red), 39 h (blue) and 138 h (green). (D), (E) and (F) GFP fluorescence detection of GFPpos in the individual fractions from the size exclusion runs of (A), (B) and (C), respectively.

Figure 4.4. Loading of intact AaLS-13 capsids. (A) Aliquots of a sample in which fractions corresponding to intact capsids were mixed with GFPpos in a ratio of 5:1 were injected on a superose 6 size exclusion column in buffer containing EDTA after 22 h (red) and 43 h (blue). (B) GFP fluorescence detection of GFPpos in the individual fractions from the size exclusion runs of (A). (C) Negative staining electron microscopy images of sample from (A). Scale bar, 100 nm.
4.3 Discussion

The experiments presented in this chapter show that the structural transitions and dynamics of AaLS-13 can be utilized to load the capsids with a positively charged protein in vitro. AaLS-13 could be loaded with about four times more GFPpos than upon coproduction in vivo. This extremely high loading suggests that in vitro encapsulation of a variety of other guests should be possible, including non-protein cargo. However, to exploit the full potential of our encapsulation system the conditions for in vitro loading will have to be further optimized to ensure reproducible and efficient capsid reassembly and loading. For quantitative reassembly it is crucial that the structure and assembly of AaLS-13 are not impaired too much during disassembly and a procedure with optimal timing has to be found. Further, the addition of LS substrate analogs to the reassembly reaction might support capsid formation. Substrate analogs have been used for the reconstitution of empty *B. subtilis* LS 60mers. After dissociation of the *B. subtilis* riboflavin-LS complex, LS forms heterogeneous assemblies of more than 100 subunits, which can be converted to T=1 icosahedral capsids by the addition of a substrate analog (23).

It should be possible to control the amount of guest encapsulated per capsid by the ratio and concentration in which host and guest are mixed for reassembly. Unfortunately, the experiment described above to test this hypothesis failed since no significant reassembly was observed. However, successful repetition of the experiment examining in vitro loading with different ratios will give insight on the dependency between the stoichiometry in the reassembly reaction and the loading.

The large amounts of encapsulated GFPpos seem to be densely packed inside the capsid as shown by the EM images. It is known that crowding and spatial confinement stabilize the folded state of proteins. AaLS-13 highly filled with GFPpos could be subjected to thermal denaturation with fluorescence detection to monitor unfolding of the encapsulated GFPpos and compared to the unfolding of GFPpos free in solution. A stabilizing effect of encapsulation in AaLS-13 would suggest possible uses of the capsid as a storage cage to keep labile proteins folded or as a folding cage to improve folding both in vivo and in vitro.
Encapsulating cargo in vitro has the advantages that loading can be more easily controlled and is specific, and that a wider range of guests can be chosen. AaLS-13 can encapsulate active enzymes, such as an esterase (see Chapter 3), and thus serve as a nanoreactor. For example, aniline can be polymerized to a special conductive polyaniline form, which is interesting for many applications including nanoelectronic devices, as material for batteries, sensors and protective coating. The chemical polymerization requires harsh conditions and often yields material with low conductivity. It has been shown that polyaniline with high conductivity can be synthesized with peroxidase enzymes under mild conditions in the presence of templates, such as negatively charged polyelectrolytes, anionic micelles, reverse micelles and surfactant vesicles (110). AaLS-13 might be able to serve as alternative template. Sequestering a positively charged isoform of horseradish peroxidase (isoform C, pI 9) should be straightforward and should allow the size-constrained synthesis of highly conductive nanomaterial inside AaLS-13, demonstration a potential use of our system for material sciences.

Encapsulation of gold nanoparticles or quantum dots (e.g. CdSe/ZnS) in AaLS-13 could be achieved by functionalization of the particles with polyethylene glycol (52) and positively charged groups. Negatively charged, citrate modified gold nanoparticles might also be loaded into AaLS-13 capsids in a metal mediated fashion as shown by encapsulation of these particles into negatively charged DNA cages (111). Caging nanoparticles, quantum dots, MRI contrast agents and other dyes in protein capsids can improve their optical properties and water solubility, and enhance their stability. Further, elements for targeting specific locations can be incorporated into the protein capsids. Protein capsids that can encapsulate such markers might be useful systems for applications in bio- and medical imaging.

In vitro loading should enable the controlled encapsulation of multiple guests and influence on their ratios inside the capsid, allowing more sophisticated challenges to be tackled. To drive the caging of two guests they could be either both positively charged or association of a positively charged guest is followed by binding of the second negatively charged guest, resulting in a defined array of the two guests. Processes where multiple components need to interact benefit from sequestration inside a capsid. Examples of such processes include enzyme reactions involving substrate channeling or regulation by effectors or inhibitors.
and electron transfer reactions, whose efficiency is strongly distance dependent. Artificial microcompartments that house whole metabolic pathways could be generated in principle by encapsulating the necessary enzymes. This would allow reactions to be performed that involve volatile or toxic intermediates and products or processes that are disturbed by other components present in the cell. Pores in the AaLS-13 capsid shell could act as a selectivity filter through which some molecules can pass whereas others are blocked based on their size and physical properties (charge, hydrophobicity). In the next chapter these pores in lumazine synthase are the target for engineering purposes.
5 Harnessing Symmetry for the De Novo Creation of Enzyme Activity

5.1 Introduction

Many protein capsules possess pores in their shell that connect luminal space with the bulk solvent. These pores act as filters that allow selective entrance and exit of desired molecules but block passage of other molecules.

Viruses exploit such pores to sense their environment and react to it. When the influenza virus is entrapped in endosomes the proton-selective M2 ion channel permits protons to enter the virus particles (112). Acidification of the lumen then causes dissociation of the capsid, a necessary step in the viral lifecycle. In the iron storage protein ferritin the gated symmetrically arranged pores control iron flow across the protein shell (25). Proteinaceous microcompartments that act as simple organelles by sequestering specific metabolic processes also contain pores that allow substrates and products to selectively pass through the protein shells (19). The carboxysome and ethanolamine utilization microcompartment are examples of the latter.

The selectivity of capsid pores provides control over the contents in the capsid interior. This control element has been successfully exploited for bio- and nanotechnological purposes. In one example, a Pd cluster was synthesized inside ferritin by in situ chemical reduction of Pd ions (69). Discriminative penetration of substrates through the pores in the protein shell conferred size selectivity on Pd-catalyzed olefin hydrogenations. The selectivity of such protein channels can also be changed by engineering their size and the polarity of the amino acids lining the tunnel. We wondered whether these pores might be redesigned to express catalytic function. Conceivably, the tunnel could be equipped with catalytic groups that would interact with appropriately sized substrates that bind in the resulting “active site”.
The engineered protein capsids from *Aquifex aeolicus* lumazine synthase described in the previous chapters encapsulate a variety of positively charged proteins and enzymes. Like other icosahedral assemblies they contain pores located at the fivefold symmetry axis of the pentameric building blocks. In the wild type enzyme they presumably allow diffusion of the natural substrates to the active site, which is located on the interior of the capsid. Indeed, the capsid cavity of *B. subtilis* LS contains the next enzyme in the pathway, riboflavin synthase, making riboflavin biosynthesis more efficient due to substrate channeling.

With an average Cα to Cα width of \(-16\) Å and a length of \(-25\) Å the LS tunnel has the appropriate dimensions for binding and catalysis of small molecules. Designing new binding or catalytic activity into a protein scaffold requires that the structure tolerate multiple clustered mutations. The tunnel in yeast LS, which does not form a capsid but assembles as a pentameric ring, has been shown to be extremely tolerant to mutations. Even extensive mutation of tunnel residues does not disrupt the protein's quaternary structure (113).

Symmetric homooligomeric proteins are potentially attractive as starting points for the creation of new activities because the structural and functional effects of a tunnel mutation in one subunit result in multiple symmetry-related changes in close proximity in the complex. The amplification of mutations due to symmetry should make symmetric ring-forming oligomers particularly sensitive to changes and thus facilitate mutagenesis strategies.

For our preliminary efforts to engineer catalytic activity into the pores of LS we have focused on pentameric yeast LS instead of the full LS capsids formed by *A. aeolicus*. The LS variant from *S. cerevisiae* (ScLS) forms pentameric structures that are nearly identical to the pentamers in the *A. aeolicus* LS capsid, and insight gained in the course of these studies should be readily extendible to capsids if successful.

In this chapter experiments to create de novo enzyme activity in the ScLS tunnel by directed evolution and rational design are described. Powerful in vivo selection systems were exploited to search for catalysts with chorismate mutase, prephenate dehydratase, and dihydrofolate reductase activities. Unfortunately, viable catalysts were not found for this set of reactions. However, rational design of catalysts for the well-understood and less demanding retro-aldol reaction afforded tunnel
variants with low aldolase activity. Directed evolution and computational design can now be used to improve catalytic activity further.

5.2 Results

5.2.1 Directed evolution with in vivo selection systems

To examine the feasibility of functionalizing the pores in lumazine synthase we initially used random mutagenesis and genetic selection for several catalytic activities. The C5 symmetric homopentamer of ScLS is assembled via extensive subunit interfaces (29% of the total accessible surface area of the monomer), which also house the LS active site (Fig. 5.1A). Lumazine synthase catalyses the penultimate step in the biosynthesis of riboflavin (Fig. 5.1B). The tunnel at the center of the ring is 12–18 Å from the active site and has no known catalytic function.

![Figure 5.1. Lumazine synthase from S. cerevisae.](image)

**Figure 5.1.** Lumazine synthase from *S. cerevisae*. (A) The pentameric structure of ScLS (82). Individual subunits are differently colored. Positions of the natural active site and the tunnel housing the potential new active site are indicated. (B) The natural reaction catalyzed by lumazine synthase.
**Library design and construction**

The tunnel of the ScLS pentameric ring is bounded by one five-turn α-helix from each subunit. Libraries were constructed by randomizing the five helix residues that project into the middle of the tunnel, namely Glu99, Asp103, His107, Asn111, Glu114 (Fig. 5.2A). The calculated water accessible surface of the protein shows that these five residues constitute most of the accessible tunnel surface (Fig. 5.2B and C). Because of symmetry, up to 25 residues are potentially changed by this procedure, providing access to enormous structural diversity starting from a relatively small library ($20^5 = 3.2 \times 10^6$).

![Figure 5.2](image)

**Figure 5.2.** The tunnel of *S. cerevisae* lumazine synthase. (A) Close-up view of the ScLS tunnel. Residues targeted for randomization are indicated. (B) and (C) Calculated water accessible surface of the tunnel from the N-terminal (B) and C-terminal (C) side. The residues targeted for randomization are displayed in green.

To create a large library of tunnel variants the codons of the five residues were randomized with a single synthetic oligonucleotide. This oligonucleotide contained the randomized codon NNB, which has the smallest probability of encoding a stop codon among the codons that allow all amino acids, at the five tunnel positions. It was incorporated into part of the LS gene by PCR (Fig. 5.3). A restriction site was introduced into the LS gene just before the randomized codons to clone the library cassette generated by PCR into the selection plasmid containing the full-length LS gene.
**Figure 5.3.** Scheme of the library construction. x represents the positions randomized by PCR with a longer primer containing the randomized codons NNB and a flanking primer symbolized by arrows. NsiI and XhoI are the restriction sites used for cloning.

**Selection for chorismate mutase activity**

Chorismate mutases (CMs) catalyze the Claisen rearrangement of chorismate to prephenate, a key step in the biosynthetic pathway leading to tyrosine and phenylalanine in plants and lower organisms. This reaction is one of the simplest enzyme-catalysed reactions and it is not particularly energetically demanding, making CM a good model system for testing ideas about catalysis, protein structure, and design. Additionally, the reaction mechanism is well understood and considerable structure-function information is available.

The importance of CM activity for bacterial growth has been exploited for the construction of a powerful CM selection system (Fig. 5.4) (114, 115). In this system, a strain of *E. coli* was engineered in which the genes encoding the bifunctional CM-prephenate dehydratase and the CM-prephenate dehydrogenase were replaced by genes encoding monofunctional versions of the dehydratase and the dehydrogenase. As a consequence of the missing CM, this selection strain is unable to grow on minimal media lacking phenylalanine and tyrosine. For growth it requires an added source of CM activity which can be provided by transformation with a plasmid encoding a functional enzyme. This system has been applied in the laboratory to test ideas about CM structure and mechanism (114, 116-118), to redesign the topology of CMs (119, 120), to apply binary patterning with a simplified amino acid alphabet to CM (121), and to create a split CM (122).
To select for LS tunnel variants with CM activity plasmids encoding the LS tunnel library were transformed into the *E. coli* CM selection strain by electroporation (121, 123). The transformed cells were streaked onto solid minimal media lacking tyrosine and incubated at 30° C. Library sizes were determined by plating serial dilutions of these cells onto LB agar, and representative clones were sequenced to verify library quality. Our relatively simple library design yielded a very high quality library: 91% of the unselected clones that were sequenced were as designed. In total, 2 x 10^7 independent clones were sampled, representing a more than six fold oversampling of the theoretical library size of 3.2 x 10^6 (= 20^5) different LS variants. However, no LS variant capable of complementing the CM deficiency of the selection strain was found, suggesting that the diverse environments created in the LS tunnel are not capable of catalyzing the CM reaction.

**Selection for prephenate dehydratase activity**

The tunnel libraries can be easily transformed into any *E. coli* selection system. While CM activity does not seem to be compatible with the LS tunnel, the tunnel might be able to catalyze other easy reactions.

Prephenate dehydratase (PDT) catalyzes the mechanistically simple dehydration and decarboxylation of prephenate to give phenylpyruvate (Fig. 5.5). The product of this reaction sequence is then transformed into phenylalanine by an aromatic
aminotransferase. Because PDT activity is also required for bacterial growth, a PDT selection system can be constructed by deleting the gene encoding natural PDT. Such systems tend to be “leaky”, however, and additional control mechanisms have been developed to regulate selection stringency (124).

![Prephenate to Phenylpyruvate Conversion](image)

**Figure 5.5.** The conversion of prephenate to phenylpyruvate catalysed by prephenate dehydratase.

Specifically, the cytoplasmic concentration of prephenate is maintained at low levels by constitutive expression of a *Z. mobilis* cyclohexadienyl dehydrogenase which converts prephenate into a precursor of tyrosine and is not subject to feedback inhibition (124). As a result, spontaneous uncatalyzed transformation of prephenate to phenylpyruvate is minimized and this selection strain is unable to grow on minimal media lacking phenylalanine.

We used this system to search for LS tunnel variants with PDT activity. The plasmids encoding the LS tunnel library were electroporated into the *E. coli* PDT selection strain (KA34 pAKZ1), and allowed to grow on minimal media in the absence of phenylalanine at 30° C. The quality of the library was again high, with 83% of the unselected clones sequenced conforming to the design. Of the 3.7 x 10⁷ individual clones tested, however, none complemented the PDT deficiency. It thus appears that the LS tunnel is incapable of supporting significant levels of PDT activity.

**Selection for dihydrofolate reductase activity**

The NADPH-dependant reduction of dihydrofolate to tetrahydrofolate, catalyzed by dihydrofolate reductases (DHFR), can be accelerated by approximation and proper orientation of cofactor and substrate (Fig. 5.6). This simplicity should convert the search for catalytic activity to the potentially simpler task of evolving a binding site.
for NADPH and dihydrofolate. This reaction is also attractive because it is known that the DHFR reaction can be catalyzed in nature by a symmetric active site. The plasmid-encoded R67 DHFR is a D₂-symmetric homotetramer, which renders bacteria resistant to the antibiotic trimethoprim (125). It provides us with a natural example of a catalyst that efficiently turns over asymmetric substrates at a symmetric active site. The X-ray structure of R67 shows that the active site is composed of symmetry related (D₂) binding surfaces from all four identical subunits and is located in the center of a 25 Å-long pore passing through the middle of the tetramer. Despite its four-fold symmetry, the tetramer only binds two substrates in an asymmetric fashion as a result of symmetry-related amino acid side chains adopting different functions. In fact, symmetry breaks are common in symmetric proteins at the atomic level. Nearly all complexes show local asymmetry at the level of side chain conformation, particularly near the symmetry axes.

**Figure 5.6.** The NADPH-dependent reduction of dihydrofolate to tetrahydrofolate catalyzed by dihydrofolate reductase.

DHFR activity is essential for cell survival as tetrahydrofolate is required for the formation of thymidylate, purine nucleosides, methionine, and other metabolic intermediates. A DHFR knockout strain in which the complete coding region of the dihydrofolate reductase gene, *folA*, was replaced with a kanamycin resistance gene (126) was used for in vivo selections of LS tunnel variants with DHFR activity. Growth of the knockout strain and the parent strain was tested on several media and at different temperatures. Conditions where the parent strain but not the DHFR knockout strain grows (minimal media supplemented with arginine at 30° C) were chosen for the selections. When the knockout strain contains plamids with *E. coli* chromosomal DNA encoding the *folA* gene for DHFR, cells grow under the selection
conditions, confirming that a plasmid-borne source of DHFR activity is readily detected by the in vivo selection system.

To identify LS tunnel variants with DHFR activity, plasmids encoding the library were electroporated into the DHFR knockout strain. This time the library quality was somewhat lower. Nevertheless, 65% of the unselected clones sequenced conformed to the design. The theoretical library size of $3.2 \times 10^6 (\approx 20^5)$ different LS variants was oversampled about four fold by testing $1.4 \times 10^7$ individual clones. Unfortunately, no LS tunnel variants with catalytic activity for the DHFR reaction were found either.

Randomizing the five LS tunnel residues per subunit followed by selection failed to yield viable catalysts for any of the three reactions tested. Since the libraries were sampled exhaustively we can be almost certain that the environments created in the symmetric LS tunnel are not capable of catalyzing the chorismate mutase, prephenate dehydratase, or dihydrofolate reductase reaction at a rate sufficient for detection by the in vivo selection systems. The chorismate mutase selection system is very sensitive for an in vivo system. It can detect catalysts with $k_{cat}/K_M$ values of about 10 M$^{-1}$s$^{-1}$ (127). The PDT selection system has been used to detect an enzyme with a $k_{cat}/K_M$ of 460 M$^{-1}$s$^{-1}$, but a less active catalyst with $k_{cat}/K_M$ of 24 M$^{-1}$s$^{-1}$ did not complement (124). However, this reaction is not very demanding chemically, requiring only the delivery of a proton to the substrate. Since completion of these experiments an improved version of the PDT selection system was developed that has tunable stringency and allows detection of less active catalysts having a $k_{cat}/K_M$ of 24 M$^{-1}$s$^{-1}$ (128). The threshold for the DHFR system has not been determined, but is probably similar to the others. Of course, weakly active LS tunnel variants in the libraries with catalytic efficiencies below the thresholds of these selection systems might have been missed.

In principle, it is possible that the tunnel of LS is simply unsuitable for the activities tested; in the case of the DHFR reaction, the available space might be too small for binding of the cofactor and the substrate. De novo creation of catalytic activity represents a challenging task because it requires the identification of constellations of properly positioned functional groups to carry out the individual steps of catalysis. In fact, there are very few successful examples in which catalytic
activity has been conferred on an initially inactive scaffold (129-132). Recent breakthroughs in computational enzyme design suggest that functional catalysts can be designed from inert protein scaffolds, at least for simple reactions (133-135). The reactions tested here might have been too demanding, they may require more extensive mutagenesis, or the detection limit for the in vivo selection systems might have been too high for clearing the hurdle of creating catalytic activity de novo. Choosing simpler reactions might be more successful.

5.2.2 Rational design of aldolase catalysts

The retro-aldol reaction represents one such simple reaction. It is mechanistically well understood, and the features required for catalysis are known. In nature, two mechanistic classes of aldolase enzymes have evolved. Class II aldolases use a mechanism requiring metal ions, usually Zn$^{2+}$. Class I aldolases, on the other hand, exploit amine catalysis with the ε-amine group of a lysine at the active site serving as a critical functional group. In the mechanism of class I aldolases the catalytic cycle is initiated by nucleophilic attack of the unprotonated lysine on the carbonyl group of the substrate (Fig. 5.7). The resulting iminium ion acts as an electron sink, facilitating deprotonation of the alcohol and concomitant C-C bond cleavage. Tautomeration of the resulting enamine to an iminium ion and subsequent hydrolysis releases the second product and regenerates the catalyst.

![Catalytic mechanism of enamine catalyzed retro-aldol reaction.](image)

**Figure 5.7.** Catalytic mechanism of enamine catalyzed retro-aldol reaction.
This mechanism has been mimicked by a variety of designed aldolases, including lysine rich α- and β-peptides (136-139), small molecules (140-143), catalytic antibodies (144), and computationally designed enzymes (134). Since imine formation is partially rate limiting in an aqueous environment, an important design feature of retro-aldol catalysts is provision of a nucleophilic lysine with a depressed pK_a that speeds up this step. The pK_a of the catalytic lysine can be lowered by placing the side chain in a hydrophobic microenvironment or through Coulombic interactions with proximal positively charge residues, such as other lysines. The first strategy has been used successfully to create catalytic antibodies and computationally designed enzymes with significant aldolase activity. The second strategy has been used to design peptide and cyclodextrin-based catalysts.

Rational design of LS tunnel variants

The tunnel of LS provides an excellent scaffold for placing multiple lysines proximal to one another and thereby perturbing their pK_a. The resulting reactive lysine could conceivable be used for the catalysis of the retro-aldol reaction of a hydrophobic aldol substrate, such as methodol (4-hydroxy-4-(6-methoxy-2-naphtyl)-2-butanone, Fig. 5.7), that binds in the tunnel. Introduction of one lysine per monomer yields a ring of five closely spaced lysines. Such a design is reminiscent of polyamino-β-cyclodextrins, which have a very similar arrangement of amines next to a binding site and were successfully used to catalyze both aldol and retro-aldol reactions (145, 146).

To design tunnel variants with potential aldolase activity, the crystal structure of ScLS was inspected. For the incorporation of the ring of lysines, residues that project into the tunnel at different depths were chosen. The lysines were introduced into the structure in pymol and positions were evaluated by placing the methodol substrate into the tunnel. Then, apolar alanines and leucines, depending on the tunnel width at this position, were selected to favor binding of the hydrophobic naphthyl group of the substrate. Based on these considerations the three simple designs depicted in Fig. 5.8 were chosen. The first design contains two rings of lysines at the N-terminal side of the tunnel, which would be the luminal side in the capsid, and a deep binding pocket for the substrate opening to the
exterior side. Design 2 has the lysines placed rather in the middle, narrow part of the tunnel, and the apolar substrate binding site would be located towards the luminal side. In the third design the lysines are positioned one helix turn further down toward the C-terminal side of the tunnel and the apolar residues reside at its narrowest position.

Figure 5.8. Rational designs of tunnel variants ScLS-1, ScLS-2 and ScLS-3. The N-terminal side of the tunnel corresponds to the luminal side in the capsid.

Retro-aldo activity of the LS tunnel variants

The designed mutations were introduced into the ScLS gene by quick change mutagenesis. The three tunnel variants M96K/E99K/D103L (ScLS-1), E99A/D103L/H107K (ScLS-2), and D103A/H107A/N111K (ScLS-3) plus wild type ScLS (ScLS-wt) were produced in E. coli and Ni²⁺-affinity purified by a C-terminal His₆-tag. All proteins were obtained in high yield (typically 60 mg to 100 mg per liter culture). CD spectroscopy revealed no changes in secondary structure in the variants compared to wild type ScLS.

The activity of the designed LS tunnel variants for retro-aldo cleavage of (±)-methodol was investigated by monitoring the formation of the fluorescent product 6-methoxy-2-naphthylaldehyde (Fig. 5,9). As test reaction conditions, 20 µM LS variant and 500 µM substrate were chosen. All variants showed detectable retro-aldo activity over the uncatalyzed reaction (Fig. 5,9). Even wild type ScLS gave almost four fold higher initial rates than the uncatalyzed background reaction, probably due to lysines on the protein surface. However, one of the designed tunnel variants, design 2 (E99A/D103L/H107K), was particularly active. Under the
conditions tested, its initial rate was 25 times faster than background and nearly seven times faster than wild type ScLS. The higher activity of this variant is presumably attributable to the mutations introduced into the tunnel, since they are the only changes relative to ScLS-wt.

Figure 5.9. Retro-aldol reaction. Retro-aldol cleavage of (±)-methodol yielding the fluorescent product 6-methoxy-2-naphthylaldehyde (blue). Table with initial velocities for cleavage of 500 µM (±)-methodol by 20 µM (pentamer) ScLS variants.

<table>
<thead>
<tr>
<th>ScLS variant</th>
<th>v [10^3 µM min^-1]</th>
<th>v[pentamer] [10^3 min^-1]</th>
</tr>
</thead>
<tbody>
<tr>
<td>ScLS-1</td>
<td>1.3</td>
<td>0.07</td>
</tr>
<tr>
<td>ScLS-2</td>
<td>7.6</td>
<td>0.39</td>
</tr>
<tr>
<td>ScLS-3</td>
<td>3.7</td>
<td>0.19</td>
</tr>
<tr>
<td>ScLS-wt</td>
<td>1.2</td>
<td>0.06</td>
</tr>
<tr>
<td>background</td>
<td>0.3</td>
<td></td>
</tr>
</tbody>
</table>

Characterization of tunnel variant ScLS-2

The most active aldolase catalyst, tunnel variant E99A/D103L/H107K, was characterized in more detail. A model of the artificial enzyme was created by fitting the mutations and the substrate into the ScLS crystal structure in pymol (Fig.5.10). As shown in Fig. 5.10 the tunnel has appropriate dimensions to bind (±)-methodol in an orientation that would place its carbonyl group in proximity to the lysine residue at position 107 (purple) so that imine formation can occur. The hydrophobic walls of the cavity (blue) would pack against the hydrophobic naphthyl group. Of the three designs, the catalytic lysine in ScLS-2 resides in the least deepest position and might therefore be more easily accessible by the substrate in a non-optimal binding site.
Figure 5.10. Model of the ScLS-2 tunnel. The mutations E99A/D103L/H107K and the (±)-methodol were fitted into the ScLS crystal structure in pymol. The lysine 107 is shown in purple, leucine 103 in blue and alanine 99 in grey.

Size exclusion chromatography was carried out to analyze the quaternary structure of tunnel variant 2 and for further purification. The mutations in the ScLS tunnel do not impair assembly of the pentameric ring, as variant 2 elutes from the column mainly at a volume corresponding to the size of a pentamer (140 ml), as does the wild type protein (Fig. 5.11). Both proteins contain small amounts of a higher assembly state, probably corresponding to capsids of 60 or 180 subunits (113 ml). This is not unprecedented, since previous experiments have shown that ScLS can form capsids in a pH and metal dependent fashion (147) and K. Woycechowsky, unpublished results). Indeed, multiple assembly states are a general feature of lumazine synthases, as demonstrated in the literature (86) and in the previous chapters for AaLS.
Because the proposed mechanism depends on a lysine residue with a perturbed pKₐ, the reaction rate should be pH dependent. Consistent with this expectation, a sigmoidal pH-rate profile is observed which plateaus above pH 8 (Fig. 5.12). The data were fitted using the equation 

\[(k_{\text{cat}}/K_M)^{\text{obs}} = (k_{\text{cat}}/K_M)^{\max}/(1 + 10^{pK_a - \text{pH}}).\]

Rates show a dependency on one ionizable group with a pKₐ of 6.8, which would be consistent with ionization of a catalytic lysine with a considerably lowered pKₐ compared to a lysine side chain in solution (~10.5). A catalytic lysine with a perturbed pKₐ value of 6.8 would be consistent with natural enzymes such as the D-2-deoxyribose-5-phosphate aldolase (DERA) in which the catalytic lysine also has a pKₐ of 6.8 (148), and with previously designed aldolases including catalytic antibodies, peptides and computationally designed enzymes having pKₐ values that range between 5.5 and 9 (136-139, 149, 150).
The catalytic efficiency of the active tunnel variant 2 was determined by steady-state kinetic measurements. The retro-aldol reaction catalyzed by design 2 follows Michaelis-Menten kinetics with a $k_{\text{cat}}$ of $1.5 \times 10^{-3}$ min$^{-1}$ and a $K_M$ of 540 µM at pH 7.5 and 30 °C (Fig. 5.13). Comparison of the catalytic rate constant of LS variant 2 with the rate constant of the uncatalyzed reaction ($6 \times 10^{-7}$ min$^{-1}$) gives a rate acceleration of $k_{\text{cat}}/k_{\text{uncat}}$ of 2500. Although substantially lower than the rate accelerations achieved by natural enzymes and catalytic antibodies, this value compares favorably with the activity of α-peptides designed and improved by phage display for this reaction. LS variant 2 has a catalytic efficiency ($k_{\text{cat}}/K_M$) that is four to nine times higher than that of evolutionary optimized α-peptide catalysts (138); its catalytic rate is also two to three times higher than the $k_{\text{cat}}$ reported for these peptides. Its turnover number is only 1.5 to 6 times lower and it is two to 16 fold less efficient than the computationally designed aldolases (134) (Fig. 5.13). However, LS variant 2 is four orders of magnitude less efficient than the catalytic antibody 38C2 (151). The binding pocket of the catalytic antibody is a very hydrophobic, 11 Å deep, slot-like cleft and is probably better suited for binding of the apolar methodol substrate than the more exposed, less hydrophobic, funnel-shaped LS tunnel (149). The catalytic lysine in the antibody is deeply buried in a hydrophobic pocket at the base of the binding site, which depresses its $pK_a$ to 6. This $pK_a$ is 0.8 units lower than of the catalytic lysine in ScLS-2 and consequently more reactive. The assistance of further catalytic groups in the pocket of the

Figure 5.12. pH-rate profile of ScLS-2. Rates were determined at 100 µM (±)-methodol and 20 µM ScLS-2 pentamer and fitted to the equation $(k_{\text{cat}}/K_M)_{\text{obs}} = (k_{\text{cat}}/K_M)_{\text{max}}/(1 + 10^{pK_a - \text{pH}})$. 

\[\text{Figure 5.12. pH-rate profile of ScLS-2. Rates were determined at 100 \mu M (\pm)-methodol and 20 \mu M ScLS-2 pentamer and fitted to the equation } (k_{\text{cat}}/K_M)_{\text{obs}} = (k_{\text{cat}}/K_M)_{\text{max}}/(1 + 10^{pK_a - \text{pH}}).\]
antibody presumably enhances catalysis of the retro-aldol reaction and accounts for the higher efficiency of the antibody compared to the LS tunnel variant. Specifically, based on the crystal structure of a related aldolase antibody bound to an enamine intermediate, it was suggested that a tyrosine residue in the active site acts as general acid and base maybe supported by other serines and an asparagine to facilitate protonation and deprotonation steps in the reaction. Mutation of the tyrosine residue to phenylalanine in the antibody 38C2 has proven its crucial role in enamine formation.

![Figure 5.13](image.png)

**Figure 5.13.** Michaelis-Menten kinetics for ScLS-2. The table shows a comparison of the catalytic parameters of ScLS-2 to other designed aldolases for the retro-aldol cleavage of (±)-methodol and to the activity of a natural aldolase. a (134); b (138); c (151); d (148).

To improve the catalytic efficiency of the rationally designed LS tunnel variant ScLS-2 directed evolution is an obvious and powerful approach. This strategy has been used very successfully to improve enzymatic properties of many different enzymes (131). The fluorescence of the retro-aldol product 6-methoxy-2-naphthylaldehyde can be used to screen for improved LS aldolases. Assays in 96 well plates have been successfully developed in our laboratory to enhance the catalytic activity of computationally designed enzymes (L. Giger, R. Blomberg, D. Hilvert, unpublished results). Preliminary results indicate that the activity of ScLS aldolases in cleared cell lysates could be similarly used to evolve improved tunnel variants. The two tunnel residues above and below the catalytic lysine, L103 and N111 are attractive sites for creating focused libraries. Residues S104, T106 and M110 would also be interesting targets since they are located near the catalytic lysine (M110) and at the potential hydrophobic binding site for the substrate (S104, T106). To keep the size of the libraries in a reasonable range for medium
throughput plate assays, reduced amino alphabets could be used (152), such as the codon NUN, which allows only hydrophobic residues F, L, I, V, M.

In preliminary experiments, the fluorogenic screening assay has been tested under different conditions of cell growth, protein expression, and cell lysis. In general, cells lysis by incubation with lysozyme and freeze-thaw cycles yielded the best results. However, under all conditions tested the activity of cell lysates containing the active LS tunnel variant ScLS-2 was close to background (Fig. 5.14A). Some differentiation between cell lysates containing variant ScLS-2 and wild type LS were seen when cell growth and protein expression were carried out in E. coli BL21 cells at 30° C, and a washing step was included, in which the cell pellets were washed with phosphate buffer to reduce background activity associated with the LB medium (Fig. 5.14B).

Figure 5.14. 96 well plate screening assay for retro-aldol cleavage of (±)-methodol. (A) Formation of the fluorescent product 6-methoxy-2-naphthylaldehyde was monitored in crude cell lysates of BL21 cells containing ScLS-wt (black) and ScLS-2 (pink), of BL21 gold cells containing no recombinant protein (green), ScLS-wt (orange), ScLS-2 (purple) and RA45.2 (blue) and wells which only contained LB medium. Cells were lysed by incubation with lysozyme and freeze-thaw cycles. (B) Product formation in crude cell lysates of BL21 cell pellets that contain ScLS-wt (black) and ScLS-2 (pink) and were additionally washed with phosphate buffer.

LS variant 2 is about ten fold more catalytic efficient (k_{cat}/K_M) than wt LS and ten fold less efficient than the aldolase RA45.2 (k_{cat}/K_M = 26 M^{-1} min^{-1}, L. Giger, unpublished results), which served as the positive control in the assays. Whereas the ten fold difference between ScLS-2 and wt is masked by the detection limit, the ten fold higher activity of RA45.2 can be readily detected in the assay. This suggests
that catalytically improved tunnel variants in libraries might also be detected by the assay.

5.3 Discussion

Although the ScLS tunnel does not appear to be able to catalyze the chorismate mutase, prephenate dehydratase, or dihydrofolate reductase reaction, it is a suitable scaffold for catalysis of the less demanding retro-aldol reaction. Simple mechanistic design principles could be easily realized in the LS tunnel due to its symmetric nature. Incorporation of a single lysine at position 107 yielded a variant with significant rate accelerations for the cleavage of the aldol substrate (±)-methanol over the uncatalyzed reaction. Usually activity is engineered into preformed pockets or active sites in proteins. Our LS variant 2 represents the first example in which catalytic activity was introduced into a protein tunnel. Its activity underscores the potential of such symmetric protein rings to serve as starting points for protein engineering.

Nevertheless, the LS catalyst ScLS-2 is not optimal and there is still a significant gap between its activity and the activities of more sophisticated enzyme models and natural enzymes. The proteinaceous and symmetric nature of the LS tunnel suggests a variety of possibilities for improving catalytic activity further.

Preliminary experiments have shown that the catalytic efficiency of LS tunnel variants might be optimized by directed evolution using a fluorogenic 96 well plate screening assay. If catalysts with a level of activity high enough to be detected by the present assay are exceedingly rare in the focused libraries, the assay can be further optimized. For example, detection could be improved by employing larger culture volumes to obtain higher enzyme concentrations. Enzyme purification steps, for example using Ni²⁺ plates, could also reduce the background. Higher throughput techniques such as in vitro compartmentalization screening systems based on microfluidics, which are currently being investigated in our group for the retro-aldol reaction, might be useful. They allow sampling of much larger libraries, which should improve the chances of finding more active catalysts.
Virtual design in silico is another possibility to facilitate the search for LS tunnels with enhanced catalytic activity. Even larger regions of sequence space can be screened using computational methods, making this an attractive option.

The symmetric nature of ScLS is a novel feature that could also be useful for generating more chemical complexity in the tunnel and increasing combinatorial diversity. Different but closely related subunits of homooligomers can assemble stochastically as homo- and heterooligomers (Fig. 5.15). Introducing a second ScLS gene into the selection vector and bicistronic expression of two distinct LS variants would allow searching for heteropentameric active sites with interesting activities (Fig. 5.15). Such sites would have increased amino acid diversity, which might extend their functional capabilities and make them more suitable for catalysis of diverse reactions. In this type of two-copy system, the combinatorial diversity in the proteome would exceed genome diversity by a factor of eight, thus extending the upper limit of library size that is experimentally manageable by almost an order of magnitude. Harnessing this diversity would provide a means of sampling sequence space more efficiently in the pursuit of novel enzymes.

Figure 5.15. Increasing chemical complexity and combinatorial diversity. The eight distinct possible combinations of two protein subunits (A, dark blue; B, light green) to form a pentameric ring are shown and the selection vector allowing bicistronic expression of two distinct LS variants.

If it were possible to make catalytically active ScLS tunnels, it might be possible to transfer the catalytic functions created to the pores of the icosahedral AaLS capsids. In this way supramolecular catalysts could be constructed. Although the two pentameric ring structures from *S. cerevisiae* and *A. aeolicus* are nearly superimposable, simple transplantation of the relevant amino acids may not be sufficient for eliciting activity in the icosahedral protein. In that case, directed
evolution with error-prone PCR or other diversification methods should rapidly adapt this scaffold to the new function.

Protein assemblies of higher order would be interesting as nanofactories that perform multiple chemical transformations in a processive manner. Natural examples include fatty acid synthase, polyketide synthase, and nonribosomal peptide synthetases, which contain multiple catalytic domains within a single protein complex to synthesize complex metabolites and natural products. Other examples are proteinaceous microcompartments like the carboxysome or ethanolamine utilization microcompartment that perform entire metabolic pathways.

Multifunctional protein particles based on LS variants with catalytically active pores could be used to encapsulate additional enzymes to create multienzyme complexes. Alternatively, capsid assembly of pentameric rings with different activities would also give rise to multifunctional assemblies. Incorporating pentamers that transfer substrates and products across the protein shell would additionally allow control over entry and exit to the capsid cavity. The various pentameric building blocks could be assembled to capsids by production in vivo, in which case fusion of the pentamers would probably be necessary to avoid interchange of monomers. It may also be possible to disassemble capsids in vitro (as in Chapter 4) and reassemble the different pentameric building blocks into a multifunctional capsid. Such a system would be highly modular and capsids could be assorted from pentamers with diverse functions as the need arises.
6 PERSPECTIVE

In this thesis the natural capsid-forming protein lumazine synthase was reengineered as a versatile nanocompartment with novel function. Designed capsid variants were optimized for efficient encapsulation of the toxic enzyme HIV protease by directed evolution. The best AaLS variant not only exhibited increased binding of HIV protease, its enhanced encapsulation capabilities could be shown to translate to other proteins. Thus, a variety of positively charged proteins can be efficiently sequestered within the lumen of the evolved capsid in vivo and in vitro, demonstrating the generality of this system and suggesting that encapsulation of other suitably charged guests, including non-protein cargo should be straightforward. Together, these results establish the capsid-forming protein lumazine synthase as an attractive nanocontainer for diverse bio- and nanotechnological applications. More specifically, they lay the foundation for developing LS-based nanocompartments that can serve as templates for the controlled synthesis of nanomaterials, or as gene and drug delivery systems. Such supramolecular systems could also be useful for bio- and medical imaging and for controlling catalysis and constructing multienzyme complexes.

As shown in Chapter 2, encapsulation of the toxic HIV protease by the evolved AaLS-13 capsid substantially increases the yield of total soluble HIV protease-R10 produced within the cells. This fact is of biotechnological relevance. Since loading of the AaLS capsids with any positively charged or R10-tagged protein should be possible, it suggests a potentially general strategy for producing appropriately tagged toxic proteins. Besides protecting the host organism from the toxic effects of the guest, AaLS capsids might also enhance the production of fragile proteins by sequestering them away from cellular proteases. Spatial confinement of proteins is believed to stabilize the folded over the unfolded state, so LS capsids might also act as folding cages, similar to chaperones, that improve folding of encapsulated guests, or as storage cages that keep labile proteins folded.

As an example of controlling nanoscale synthesis, LS capsids could be used as templates to polymerize aniline. A special electrically conductive form of polyaniline, the emeraldine salt, is interesting for many applications, including
nanoelectronic devices, batteries, sensors, and protective coatings, because of its low cost, high stability, and good electronic and optical properties. However, chemical polymerization often yields material with low conductivity and requires harsh conditions. In the presence of templates, polyaniline with high conductivity can be synthesized with peroxidase enzymes under mild conditions. As templates, negatively charged polyelectrolytes, anionic micelles, reverse micelles and surfactant vesicles (110) have been used. AaLS-13 capsids might represent a useful alternative. The selective encapsulation of a positively charged isoform of horseradish peroxidase (such as isoform C with a pl of ca. 9, or isoforms D or E, which also have high pl values) should be straightforward. The aniline monomers should be able to pass through the tunnels in the capsid walls and enter the lumen where the encapsulated horseradish peroxidase would catalyze their polymerization to monodisperse and highly conductive nanomaterial.

Encapsulation of cargo with useful optical and magnetic properties could lead to the development of host-guest complexes for bio- and medical imaging. Potentially interesting guests include the highly fluorescent GFP, which is very efficiently encapsulated by AaLS-13, but also gold nanoparticles, quantum dots (e.g. CdSe/ZnS), MRI contrast agents or other dyes. The protein shell can improve the optical and magnetic properties and stability of such cargo, and it should increase biocompatibility and solubility, while reducing toxicity of the encapsulated compounds. To tap the full potential of LS capsids for in vivo imaging purposes it should be possible to further modify the capsid exterior by genetic or chemical methods to target specific cells or tissues. For example, tumor targeting peptides, cell internalization sequences to promote cell entry, and polyethylene glycol to lower immune response could be attached at specific surface sites. Such exterior modification is likely to be crucial for the development of drug and gene delivery systems.

The strategies outlined in this thesis should be easily extended to the encapsulation of medicinally relevant compounds or nucleic acids. Towards the latter goal, the charges of the AaLS-13 encapsulation system would have to be inverted. Preliminary results in our laboratory indicate that AaLS capsids possessing a positively charged interior surface do indeed encapsulate nucleic acids efficiently. Attaching cell internalization tags to these AaLS capsids is in
progress and could be extended to append sequences that target the capsids to specific cell types or tissues.

As seen in nature, colocalizing multiple catalysts is beneficial with respect to the efficiency and control of catalysis. AaLS protein capsids could serve as the basis for the construction of artificial multienzyme complexes that function as nanofactories performing multiple chemical transformations in a processive manner. Organization of catalysts in such complexes should enhance the efficiency of the overall process due to channeling of substrates and improving energy transfer reactions, which are strongly distance dependent. Sequestration allows processes that involve volatile or toxic intermediates to be performed and it can prevent side reactions and inhibition. As a proof of principle, artificial carboxysomes (compartments that enhance CO$_2$ fixation) might be engineered by encapsulating appropriately charged or tagged versions of the enzymes carbonic anhydrase and RubisCO in AaLS-13. A more challenging goal of high interest would be the construction of an artificial photosynthetic system by sequestering catalysts that promote the conversion of sunlight energy to the synthesis of useful products. Colocalization and organization of the different catalysts that perform these energy transfer processes in complexes is crucial for efficient reaction. Such photosynthesis mimics would be of considerable interest for environmentally friendly, energy saving chemical syntheses and the retrieval of renewable energies.

LS capsids possess pores that connect the capsid interior to the bulk surrounding. Building on the work presented in Chapter 5 showing that these sites can be endowed with novel chemical function, engineering these tunnels would provide a means of controlling entry and exit of molecules across the capsid shell. They could also facilitate the construction of multienzyme complexes by serving as additional catalytic sites.

Although only a few ideas for possible future applications of AaLS based encapsulation systems are mentioned here, the robustness of LS structures to reengineering and the apparent ease with which they could be endowed with new function suggest that these protein scaffolds will be excellent starting points for many exciting applications. This system beautifully illustrates the power of combining rational design with Darwinian evolution for the field of nanotechnology.
7 MATERIALS AND METHODS

7.1 Materials and standard methods

Buffers and salts were purchased from commercial suppliers. Oligonucleotides were custom-synthesized and purified by MicroSynth. Restriction enzymes and Taq polymerase were from New England BioLabs. Murine monoclonal anti-HIV protease antibody was obtained from Exbio, while goat anti-mouse IgG antibody conjugated to horse radish peroxidase was obtained from Sigma-Aldrich. DNA sequencing was carried out on a 3100-Avant Genetic Analyzer (Applied Biosystems) by chain termination chemistry, using the BigDye Terminator Cycle Sequencing Kit from the same company. Wild type HIV protease was a generous gift of Celia Schiffer. (±)-methodol ((±)-(4-hydroxy-4-(6-methoxy-2-naphtyl)-2-butanone) was synthesized by Lars Giger according to (151). Plasmid pET-GFP-POS36 (101) encoding the supercharged GFP variant was a generous gift of David R. Liu. Other microbiological and molecular biological methods were performed according to standard procedures (153).

7.2 Plasmid construction

pACYC-tetHIV-R10

The plasmid pACYC-tetHIV-R10 was used for the tetracycline-inducible production of the C-terminally R10-tagged HIV protease and was constructed by subcloning the gene encoding HIV protease, originally stemming from plasmid pET9cPrt (154), into a pACYC184 (155) derived vector that contains the tetracycline promoter (156). To assemble pACYC-tetHIV-R10 (3569 bp), the HIV protease gene was excised from vector pMG-HIV-GFPH6 (Martin Neuenschwander, unpublished results) using NdeI and Xhol restriction enzymes. The fragment (302 bp) was then ligated into vector pAC-Ptet-GFP (3267 bp, (157)), which had been digested with the same enzymes to remove the GFP gene.
**pACYC-GFPpos and pACYC-GFPpos-R10**

The gene encoding the supercharged variant GFPpos was amplified from vector pET-GFP-POS36 (101) by overlap extension PCR, which at the same time removed an internal NdeI restriction site. Primers GFPpos36fwd (TTACGTCATATGGCTAGCAAAGGTGAACGTC), GFPpos36-Nde-a (ATCGTGACGTTTTCATGTGTATGGGTAACGAC), GFPpos36rev (TGCTATCTCGAGCGCCATGGAGCCGCCGACTTGTAGCCTCGTCG), GFPpos36-Nde-s (GTCGTTACCCTAAACATGAAACGTCACGAT) were used for this overlap extension PCR. The NdeI and XhoI digested PCR fragment was ligated into vector pAC4C-GFP and pAC4C-GFP-R (81, 147), which had been digested with the same enzymes to remove the GFP gene. The resulting plasmids pACYC-GFPpos and pACYC-GFPpos-R10 encode GFPpos and GFPpos-R10 under control of the T7 promoter.

**pMG-ScLS-wt, pMG-ScLS-1, pMG-ScLS-2, pMG-ScLS-3**

pMG-ScLS-wt was obtained by introducing a silent mutation into the ScLS gene on a pMG209 derived plasmid (113) using quick change mutagenesis with primers bw2 (TAGTACAATGCATTTTGAATACATT) and bw3 (AATGTATTCAAAATGCAATTGTACTA) to create a NsiI restriction site that allows cloning of the tunnel library cassette. Mutations of the aldolase designs were introduced into the ScLS gene by quick change mutagenesis using the following primers:

- D103Ls (CACTTTAAATACATTTCATTATCCACTACTCACG) and
- D103La (CGTGAGTGTGGATACATTTCATTATCCACTACTACG) resulting in pMG-ScLS-1,
- E99As (GGTAGTACAATGCACTTTGCATACATTTCACTTTCCACTACTAAGGCATTGATGAAATTACA) and
- E99Aa (TTGTAAGTTCATCAATGACTTTGCATACATTTCACTTTCCACTACTAAGGCATTGATGAAATTACA) resulting in pMG-ScLS-2,
- H107As (CACTTTGAATACCATTTGCATACATTTCACTTTCCACTACTAAGGCATTGATGAAATTACA) and
7.3 Library construction

Library construction of ScLS tunnel variants

The five tunnel residues Glu99, Asp103, His107, Asn111, Glu114 of ScLS were randomized in a standard PCR using primer bw1 (TAGTACAATGCATTTTNNBTACATTTCCANNBTCCACACTNNBGCATTGATG)

NNBTACAAANNNBAAAGTCCGACATGCGCTG) and 140104YLSa (GATATACTCGAGAAAAGCATTTTTACCGAACT) and the plasmid pMG-ScLS-wt as template. The resulting PCR product was purified by agarose gel electrophoresis. The gel-purified (NuceloSpin Extract II Kit, Macherey-Nagel) PCR product and the acceptor vector pMG-ScLS-wt were digested with Nsil and XhoI. The digested PCR product was purified with the JetQuick PCR purification kit (GENOMED, Chemie Brunschwig AG), and the digested vector fragment was purified by agarose gel electrophoresis and extraction. The fragments were ligated with T4 DNA ligase to give a plasmid library containing the mutant ScLS genes. After phenol/chloroform extraction, the ligation product was desalted and concentrated using a Microcon YM-30 filter (Millipore) and 2 mM Tris-HCl (pH 8) as wash buffer.

Error-prone PCR library

To generate a large collection of mutants derived from AaLS-neg, the plasmid pMG-AaLS-neg (81, 117) was used as a template in an error-prone PCR (epPCR) reaction, which was carried out using the GeneMorph II Random Mutagenesis Kit from Stratagene according to the manufacturer’s instructions aiming at a medium to high mutation rate (4.5 to 16 mutations/kb). The primers AQs (GATATACCATGGAAATCTACGAAGTAAACTA) and AQa (GATATACTCGAGTAGGAGACTTGAATAAGT), flanking the coding region of the capsid gene, were employed for amplification. The resulting epPCR product
(482 bp) was purified by agarose gel electrophoresis. The gel-purified (NuceloSpin Extract II Kit, Macherey-Nagel) epPCR product and the acceptor vector pMG-AaLS-neg were digested with Ncol and XhoI. The digested PCR product (464 bp) was purified with the JetQuick PCR purification kit (GENOMED, Chemie Brunschwig AG), and the digested vector fragment was purified by agarose gel electrophoresis and extraction. The fragments were ligated with T4 DNA ligase to give a plasmid library containing the mutant capsid genes. After phenol/chloroform extraction, the ligation product was desalted and concentrated using a Microcon YM-30 filter (Millipore) and 2 mM Tris-HCl (pH 8) as wash buffer.

**DNA shuffling library**

For the second and third rounds of evolution, genes encoding capsid variants selected in the previous round were subjected to DNA shuffling essentially as published (158). Briefly, capsid genes were amplified from the isolated plasmids in an initial PCR reaction using primers AQs and AQa. The PCR product (482 bp) was gel-purified, and ca. 3 µg were digested for about 75 s with DNaseI (6 ng/µl) to give 50 bp to 100 bp fragments in 50 µl 20 mM Tris-HCl, pH 8, 10 mM MgCl₂. Digestion was halted by shock freezing in liquid nitrogen, addition of EDTA to a final concentration of 50 mM, followed by heat inactivation at 65 °C for 10 min. The fragments were purified by agarose gel electrophoresis, concentrated using a Microcon YM-30 filter, and assembled in a PCR-like process, but without primers, at concentrations of 5-10 ng/µl. The reassembled genes were amplified in a final PCR using primers AQs and AQa, digested with Ncol and XhoI, and ligated into the pMG vector, desalted and concentrated as described in the previous section to afford plasmid libraries containing mutant capsid genes.

**7.4 Selections**

**Selections for CM activity**

Plasmids encoding ScLS tunnel variants (about 1 µg plasmid per 50 µl competent cells) were transformed into electrocompetent *E. coli* KA12 pKIMP-UAUC (Kast 1996 PNAS) and then incubated in 15 ml SOC medium per 50 µl competent cells for
50 min at 30 °C. Cells were washed three times with liquid M9 minimal medium, and aliquots were plated on selective M9c+F minimal plates (117) containing 20 µg/ml L-phenylalanine, 150 µg/ml ampicillin, 30 µg/ml chloramphenicol, and 100 µM salicylate to induce ScLS variant production, and incubated at 30 °C. The library size was determined by plating serial dilutions of these cells onto M9c+F+Y plates containing 20 µg/ml L-tyrosine. Representative clones were sequenced to verify library quality.

**Selections for PDT activity**

Plasmids encoding ScLS tunnel variants (about 1 µg plasmid per 50 µl competent cells) were transformed into electrocompetent *E. coli* KA34 pAKZ1 (Diss Andreas) and then incubated in 15 ml SOC medium per 50 µl competent cells for 50 min at 30 °C. Cells were washed three times with liquid M9 salt, and 50 µl aliquots were plated on selective M9c+Y minimal plates (117) containing 20 µg/ml L-tyrosine, 150 µg/ml ampicillin, 30 µg/ml chloramphenicol, and 100 µM salicylate to induce ScLS variant production, and incubated at 30 °C. The library size was determined by plating serial dilutions of these cells onto M9c+F+Y plates containing 20 µg/ml L-phenylalanine. Representative clones were sequenced to verify library quality.

**Selections for DHFR activity**

Plasmids encoding ScLS tunnel variants (about 1 µg plasmid per 50 µl competent cells) were transformed into electrocompetent *E. coli* MH829 (126) and then incubated in 15 ml SOC medium per 50 µl competent cells for 50 min at 30 °C. Cells were washed three times with liquid M9 minimal medium, and aliquots were plated on selective M9+R minimal plates containing M9 media (M9 salts, 0.2% (w/v) D-glucose, 1 mM MgSO₄, 0.1 mM CaCl₂), and 5 µg/mL thiamine-HCl, 50 µg/ml L-arginine, 150 µg/ml ampicillin, 30 µg/ml kanamycin, and 100 µM salicylate to induce ScLS variant production, and incubated at 30 °C. The library size was determined by plating serial dilutions of these cells onto LB plates containing 50 µg/ml thymidine, 150 µg/ml ampicillin, 30 µg/ml kanamycin, and 100 µM salicylate. Representative clones were sequenced to verify library quality.
Selection for improved AaLS capsid variants

Plasmids encoding capsid variants with enhanced protease encapsulation ability were selected by growing pools of cells containing individual members of the plasmid libraries under conditions of high protease production. Electrocompetent E. coli XL1-Blue (159) cells were transformed with the capsid libraries (ca. 2.5 μg plasmid) by electroporation and then incubated in 30 ml SOC medium for 50 min at 30 °C. The library size was determined by plating of serial dilutions of these cells onto LB-agar containing ampicillin (150 μg/ml), and representative clones were sequenced to verify library quality. To the remaining cells, SOB medium and ampicillin (150 μg/ml) were added to a final volume of 50 ml. These precultures were grown overnight at 30 °C and then used to inoculate cultures for the preparation of electrocompetent cells (153) containing the capsid library, to enable transformation with plasmid pACYC-tetHIV-R10, encoding the R10-tagged HIV protease. After electroporation, the cells were incubated in 45 ml LB medium for 50 min at 30 °C before adding ampicillin (150 μg/ml) and chloramphenicol (30 μg/ml). This two-step transformation procedure was necessary for each round of selection to preclude accumulation of undesired mutations in the protease gene that would negate the selection pressure.

Upon addition of the antibiotics, capsid production was induced by the addition of salicylate to a final concentration of 100 μM followed by a 1 h incubation at 30 °C and 250 rpm. To begin the selection, the culture containing the capsid library was split into 4 ml portions and the HIV protease-R10 gene was induced with different concentrations of tetracycline, ranging from 400 to 1600 ng/ml (round I and II: 400, 600, 800, 1000 and 1200 ng/ml; round III and IV: 600, 800, 1000, 1200, 1400 and 1600 ng/ml). The cells were then grown at 30 °C and 250 rpm for 28 h in round I, for 26 h in rounds II and III, and for 24 h in round IV. As a positive control for growth rates, a 4 ml culture of cells producing the capsid library and containing the plasmid encoding HIV protease-R10 was incubated in parallel without tetracycline. As internal standards and negative controls, cells containing plasmids encoding AaLS-wt or AaLS-neg, rather than the capsid library, were transformed with pACYC-tetHIV-R10 and treated as the cells containing the library. After incubation, cells were harvested from the cultures containing the library and plasmids encoding the surviving capsid variants were isolated (JETquick Plasmid
Miniprep Kit, GENOMED, Chemie Brunschwig AG). Because sequencing of representative variants obtained under the different selection conditions revealed no obvious trends towards a consensus sequence, the plasmids from cultures grown at or above 600 ng/ml tetracycline were pooled for the next round of diversification and selection. For the fourth round, capsid genes were amplified from plasmids selected in the third round without further diversification by PCR (primers AQs and AQa) and recloned into fresh acceptor vector again using restriction sites Ncol and XhoI as described in the section describing construction of the error-prone PCR library. After round IV, 24 variants were isolated, sequenced, and characterized.

7.5 Growth tests

Plots of cell density vs. time were generated to assess the ability of individual capsid variants to protect cells from the harmful effects of intracellular R10-tagged HIV protease. E. coli XL1-Blue cells were transformed with plasmids pMG-AaLS-wt, pMG-AaLS-neg or pMG-AaLS-13, which encode the different capsid variants, and pACYC-tetHIV-R10, which encodes HIV protease-R10. Single colonies were used to inoculate precultures in LB medium (5 ml) supplemented with ampicillin (150 μg/ml), chloramphenicol (30 μg/ml), and salicylate (100 μM), which induces production of the AaLS capsid. These precultures were incubated overnight at 30 °C and 250 rpm and then used to inoculate a second batch of precultures in fresh LB medium (4 ml) containing ampicillin (150 μg/ml), chloramphenicol (30 μg/ml), and salicylate (100 μM) at a starting OD600 of 0.05. These second precultures were grown to an OD600 of ca. 0.5 under the same conditions as the first precultures, and used to inoculate 50 ml LB containing ampicillin (150 μg/ml), chloramphenicol (30 μg/ml), and salicylate (100 μM) at a starting OD of 0.01. This 50 ml culture was split into 4 ml portions and production of HIV protease-R10 was induced with different tetracycline concentrations. The cultures were incubated at 30 °C and 250 rpm for 28 h, and the OD600 was recorded every hour.
7.6 Monitoring the stress response by real-time PCR

_E. coli_ XL1-Blue cells were transformed with plasmids pMG-AaLS-13 and/or pACYC-tetHIV-R10. Single colonies were used to inoculate precultures in LB medium (5 ml) supplemented with ampicillin (150 μg/ml) and/or chloramphenicol (30 μg/ml) and incubated overnight at 30 °C and 250 rpm. These precultures were then used to inoculate cultures in fresh LB medium (4 ml) containing ampicillin (150 μg/ml) and/or chloramphenicol (30 μg/ml) at a starting OD600 of 0.05. The cultures were grown under the same conditions as the precultures. At an OD<sub>600</sub> of ca. 0.5 production of AaLS was induced with salicylate (0.1 mM); production of HIV protease was induced with tetracycline (700 ng/ml). For the heat shock control XL1-Blue cells were grown as described above and at an OD600 of ca. 0.5 900 μl of culture was transferred to 50 °C for 7 min. Aliquots of 900 μl culture were removed before induction or heat shock, and after 7 min, 15 min and 30 min. RNA in the samples was immediately stabilized using RNAprotect Bacteria Reagent (Qiagen), and total RNA was isolated using RNeasy Mini Kit (Qiagen) according to the manufacturer’s instructions. DNA was eliminated by an on-column digest with RNase-free DNase (Qiagen) during the RNA isolation step, followed by a second digest in solution and RNA cleanup using the RNeasy Mini Kit. To ensure that genomic DNA was completely removed, a PCR using primers rpoD_for and rpoD_rev (see below) for the rpoD gene was carried out for 30 cycles. 1 μg of total RNA was reverse transcribed with Omniscript reverse transcriptase (Qiagen) using random nonamers. Quantitative real-time PCR reactions for the groEL, dnaK, ibpA and rpoD genes were performed on a Rotor-Gene 3000 instrument (Corbett Life Science) with Rotor-Gene Probe PCR Kit (Qiagen) according to the manufacturer’s instructions using the primers and the FAM and BHQ-1 labeled probes listed below. Relative expression ratios for groEL, dnaK and ibpA were calculated from the efficiencies of the PCR reactions and the crossing point (CP) deviations between samples before and after induction or heat shock. The values were normalized using the gene rpoD as an endogenous control (160).
Primers and probes used for real time PCR:

<table>
<thead>
<tr>
<th>Primers</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>groEL_for</td>
<td>TATCAAAGTGCTGCTGCTACC</td>
</tr>
<tr>
<td>groEL_rev</td>
<td>AGTTTAGACGCTACGCGGATC</td>
</tr>
<tr>
<td>groEL_probe</td>
<td>FAM-AACACCACCAGCAACCAGGCC-BHQ-1</td>
</tr>
<tr>
<td>dnaK_for</td>
<td>CGAACCGCAAAAACACTCTG</td>
</tr>
<tr>
<td>dnaK_rev</td>
<td>ATGGGAAACATCACGCTGACTTC</td>
</tr>
<tr>
<td>dnaK_probe</td>
<td>FAM-AACGCCTGATTGTCGCGCTTCCC-BHQ-1</td>
</tr>
<tr>
<td>ibpA_for</td>
<td>CAGAGTAATGGCGGCTACCC</td>
</tr>
<tr>
<td>ibpA_rev</td>
<td>GGTAATTTCCAGTTCGCTCTCAG</td>
</tr>
<tr>
<td>ibpA_probe</td>
<td>FAM-CCACGGCAGCGATGACGCGG-BHQ-1</td>
</tr>
<tr>
<td>rpoD_for</td>
<td>GTGACACCATCAAAGCAGAAAGG</td>
</tr>
<tr>
<td>rpoD_rev</td>
<td>ACCAGGTAGTCAAAGCTGCTTCG</td>
</tr>
<tr>
<td>rpoD_probe</td>
<td>FAM-CGCAGTCAGCTACCAGGAA-BHQ-1</td>
</tr>
</tbody>
</table>

7.7 Protein production and purification

Protein production and purification of ScLS variants

ScLS variants were overproduced using the T7 promoter system (117) in E. coli BL21 cells (161) that had been transformed with the appropriate plasmid (pMG-ScLS-wt, pMG-ScLS-1, pMG-ScLS-2, or pMG-ScLS-3). Cultures were grown at 37 °C in LB medium (400 ml) supplemented with ampicillin (150 μg/ml) to an OD₆₀₀ of 0.7 and then production of AaLS was induced with IPTG (0.1 mM). Four h after induction, the cells were harvested by centrifugation at 4200 g and 4°C for 10 min. The cell pellets were resuspended in 8 ml of lysis buffer (50 mM sodium phosphate, 300 mM NaCl, pH 8.0) and lysed by incubation with lysozyme (0.5 mg/ml) for 1 h at room temperature followed by sonication. Cell lysates were cleared by centrifugation (10,000 g for 25 min at 20 °C) and loaded onto 2.5 ml Ni²⁺-NTA agarose resin (Qiagen) in a gravity flow column that had been equilibrated with lysis buffer. After washing with lysis buffer and lysis buffer containing 40 mM imidazole, ScLS variants were eluted with lysis buffer containing 500 mM imidazole. Proteins were dialyzed into buffer containing 50 mM sodium phosphate, 150 mM NaCl, pH 7.5. Protein purity was assessed by SDS-PAGE. Protein concentrations were determined using the Coomassie Plus Protein Assay Reagent (Pierce) with bovine serum albumin (BSA) as the standard.
ScLS variants were further purified for Michaelis-Menten kinetics and pH rate profile measurements by size exclusion chromatography on a Superdex 75 HiLoad FPLC column (Amersham Pharmacia) using 50 mM sodium phosphate, pH 7.5, containing 150 mM NaCl as running buffer.

**Protein production and purification of AaLS variants and coproduction of HIV protease**

For detailed biophysical characterization, AaLS variants were overproduced using the T7 promoter system (117) in *E. coli* KA13 cells (162, 163) that had been transformed with the appropriate plasmid (pMG-AaLS-wt, pMG-AaLS-neg, or pMG-AaLS-13). For coproduction with HIV protease, the cells were also transformed with either pACYC-tetHIV or pACYC-tetHIV-R10. Cultures were grown at 30 °C and 250 rpm in LB medium (100 ml) supplemented with ampicillin (150 μg/ml) and, for coproduction with HIV protease, chloramphenicol (30 μg/ml). At an OD_{600} of 0.7, production of AaLS was induced with IPTG (0.1 mM). For coproduction of HIV protease, tetracycline (5000 ng/ml) was added immediately afterwards. 20 h after induction, the cells were harvested by centrifugation at 4200 g and 4°C for 10 min. The cell pellets were resuspended in 2 ml of lysis buffer (50 mM sodium phosphate, 300 mM NaCl, 10 mM imidazole, pH 8.0) and lysed by incubation with lysozyme (0.5 mg/ml) for 1 h at room temperature followed by sonication. Cell lysates were cleared by centrifugation (10,000 g for 25 min at 20 °C) and loaded onto 1 ml Ni^{2+}-NTA agarose resin (Qiagen) in a gravity flow column or onto Ni^{2+}-NTA spin columns (Qiagen) that had been equilibrated with lysis buffer. After washing with lysis buffer and lysis buffer containing 40 mM imidazole, capsids were eluted with lysis buffer containing 500 mM imidazole. EDTA was added immediately to the eluted fractions at a final concentration of 5 mM. Protein purity was assessed by SDS-PAGE. Protein concentrations were determined using the Coomassie Plus Protein Assay Reagent (Pierce) with bovine serum albumin (BSA) as the standard.
Protein production and purification of AaLS variants and coproduction of GFP variants

When AaLS variants were coproduced with GFP or GFP-R_{10}, the T7 promoter system was used for overexpression of the capsid variants and the GFP variants in *E. coli* BL21 cells. The cells were transformed with plasmids pMG-AaLS-neg or pMG-AaLS-13, which encode the different capsid variants, and pAC4C-GFP or pAC4C-GFP-R (81, 147), which encode the untagged and tagged GFP, respectively. Cultures were grown at 30 °C and 250 rpm in LB medium (100 ml) supplemented with ampicillin (150 μg/ml) and chloramphenicol (30 μg/ml). At an OD$_{600}$ of 0.7, production of AaLS and GFP was induced with IPTG (0.1 mM). 20 h after induction, the cells were harvested by centrifugation at 4200 g and 4°C for 10 min. Proteins were purified as described in the previous section.

For coproduction of AaLS variants with the supercharged GFPpos or GFPpos-R$_{10}$, proteins were also overexpressed using the T7 promoter system in *E. coli* BL21 cells. The cells were transformed with plasmids encoding the different capsid variants (pMG-AaLS-neg or pMG-AaLS-13) and pACYC-GFPpos or pACYC-GFPpos-R$_{10}$, which encode the untagged and tagged supercharged GFP, respectively. The recovery and growth of the cells after transformation was carried out at room temperature. Proteins were produced as described above for GFP, but the cells were grown at 25 °C instead of 30 °C at all steps and the cultures were harvested about 24 h after induction.

Protein production and purification of GFPpos

The supercharged GFP variant GFPpos was produced in *E. coli* BL21 cells, basically as described in the previous section. Cells that had been transformed with plasmid pACYC-GFPpos were grown at 25 °C and 250 rpm in LB medium (400 ml) supplemented with ampicillin (150 μg/ml). At an OD$_{600}$ of 0.7, production of GFPpos was induced with IPTG (0.1 mM). About 24 h after induction, the cells were harvested by centrifugation at 4200 g and 4°C for 10 min. The cell pellets were resuspended in 7 ml of buffer A (50 mM sodium phosphate, pH 8.0) and lysed by incubation with lysozyme (0.5 mg/ml) for 1 h at room temperature followed by sonication. Cell lysates were cleared by centrifugation (10,000 g for 25 min at
20 °C) and purified by cation exchange chromatography (Mono S HR 10/10, Pharmacia) with a 300 ml linear gradient of buffer A to buffer B (50 mM sodium phosphate, 2 M NaCl, pH 8.0). The GFPpos-containing fractions were pooled, concentrated and washed once with buffer A to lower the NaCl concentration to 100-200 mM using a Microcon YM-10 filter (Millipore). Protein purity was assessed by SDS-PAGE. Protein concentrations were determined using the Coomassie Plus Protein Assay Reagent (Pierce) with bovine serum albumin (BSA) as the standard.

7.8 Detection of HIV protease

ELISA measurements

Ni²⁺-NTA affinity purified capsid samples were denatured in SDS gel loading buffer at 80 °C for 20 min and then coated onto Immunomaxi 96 well plates (TPP) in dilutions of 3-fold, 10-fold and 30-fold. Copurifying HIV protease was detected with a commercial murine anti-HIV protease antibody, at a dilution of 1:3000, as the primary antibody and a horseradish peroxidase-conjugated anti-mouse IgG antibody, diluted 1:10000, as the secondary antibody. The plates were developed with the horseradish peroxidase substrate 3,3′,5,5′-tetramethylbenzidine (Sigma), which forms a blue reaction product, that was detected by absorption at 370 nm.

Western Blot analysis

For Western blot analysis, protein production was carried out essentially as described in the “Protein production and purification” section. For the samples used to generate Fig. 2.5, protein production was carried out using 200 ml cultures grown at 37 °C and harvested four h after induction. HIV protease was produced in the absence of capsid by first growing KA13 cells harboring either the pACYC-tetHIV or pACYC-tetHIV-R10 plasmid at 37 °C in LB medium (100 ml) supplemented with chloramphenicol (30 μg/ml) to an OD₆₀₀ of about 1.2, and then adding tetracycline (5000 ng/ml). Cells were harvested after incubation for an additional 4 h at 37 °C.
Cleared cell lysates were subjected to Ni\textsuperscript{2+}-NTA affinity chromatography. The protein concentrations in the 500 mM imidazole fractions were then determined with the Coomassie Plus Protein Assay Reagent. AaLS-containing samples were adjusted to have the same total protein concentration by dilution of the more concentrated sample. For the samples used to generate Fig. 2E additional dilutions were prepared as indicated. Samples lacking AaLS contained undetectable amounts of protein in the 500 mM imidazole fractions and were not diluted further. SDS-PAGE was carried out on 20% acrylamide gels using the Phast System (GE Healthcare), and was followed by blotting onto nitrocellulose membranes. The AaLS variants were visualized by Coomassie Blue staining of the SDS-PAGE gels. HIV protease on the nitrocellulose membranes was detected using a commercial murine anti-HIV protease antibody at a dilution of 1:3000 as the primary antibody and a horseradish peroxidase-conjugated anti-mouse IgG antibody diluted 1:10000 as the secondary antibody. Blots were developed with the luminol and peroxide-based substrate Immobilon Western (Millipore). Chemiluminescence was detected on Amersham Hyperfilm ECL (GE Healthcare).

7.9 Mixing experiments

AaLS capsids and HIV protease-R\textsubscript{10} were each produced in separate cultures of \textit{E. coli} KA13 cells using the conditions described in the sections on “Protein production and purification” and “Detection of HIV protease by Western Blot”. Cell pellets from a 50 ml culture of cells producing AaLS-13 and a 100 ml culture of cells producing HIV protease-R\textsubscript{10} were each resuspended in 1 ml lysis buffer (50 mM sodium phosphate, 300 mM NaCl, 10 mM imidazole, pH 8.0) and combined. The pooled cells were lysed by incubation with lysozyme (2 mg) for 1 h at room temperature and then subjected to sonication. After centrifugation at 10,000 \(g\) and 20 °C for 25 min, lysis buffer was added to the cleared cell lysate to a final volume of 3.5 ml, and the diluted lysate was then split into two halves. To the first half, lysis buffer containing 5 M NaCl was added to give a final salt concentration of 1 M. An identical volume of lysis buffer containing 300 mM NaCl (i.e., the same lysis buffer
used for cell lysis) was added to the other half of the lysate. Following a 15 min incubation at room temperature, the samples were each loaded onto Ni²⁺-NTA agarose resin (0.25 ml) in gravity flow columns that had been pre-equilibrated with lysis buffer containing the same salt concentration as the protein sample (1 M NaCl or 300 mM NaCl). After loading, the column was washed with 12.5 ml lysis buffer (keeping the salt concentration constant), followed by 12.5 ml lysis buffer containing 300 mM NaCl for both samples, and 5 ml lysis buffer containing 40 mM imidazole and 300 mM NaCl. The protein was then eluted with lysis buffer containing 500 mM imidazole and 300 mM NaCl in 250 μl fractions. EDTA was added immediately to all eluted fractions at a final concentration of 5 mM. The concentration of the purified proteins was determined with the Coomassie Plus Protein Assay Reagent, and the samples were analyzed by SDS-PAGE and Western blotting as described above. For comparison, the effects of a high salt concentration were also tested on the complex formed upon coproduction of AaLS-13 and the R₁₀-tagged HIV protease. In this case, the cell pellet resulting from a 50 ml culture was resuspended in lysis buffer (6 ml) and the cleared cell lysate (6.3 ml) was then split into two halves, which were incubated in either 1.0 M NaCl or 300 mM NaCl and purified in the same way as described above for the samples obtained by in vitro mixing of lysates.

7.10 Electron microscopy (EM)

Negative staining EM

Purified capsids (0.5 mg/ml to 2 mg/ml) were absorbed onto glow discharged EM grids by positioning the grid onto a droplet of sample solution for 1 min. The grids were rinsed twice with distilled water. After each rinsing step the grids were blotted dry with filter paper. The negative staining was carried out using 2 % (w/v) aqueous uranyl acetate (pH 4) and consisted of a 2-3 s rinse (and blotting) followed by a 15-30 s incubation with the stain solution. The samples were visualized with a Philips CM12 microscope (100kV equipped with a Gatan CCD camera model 794). Diameters of the capsids in the images were measured using the software ImageJ.
(http://rsb.info.nih.gov/ij/) and are reported as the average of 497 particles for AaLS-neg, 252 particles for AaLS-13 and 629 particles for AaLS-13 coproduced with HIV protease-R10.

**Cryo-transmission electron microscopy (Cryo-EM) and data analysis**

Protein sample (3.5 µl of a 2 mg/ml solution) containing purified AaLS-13 (either coproduced with HIV protease-R10 or not) was mounted onto holey carbon grids (Quantifoil), blotted to make thin aqueous films under controlled room temperature and humidity conditions (~95%), and plunged into liquid ethane using a Vitrobot apparatus (FEI Company). The grids were examined at the temperature of liquid nitrogen using a cryo-holder (model 626, Gatan) and a Tecnai G2 F20 microscope (FEI Company) equipped with a field emission gun and energy filter (Gatan) operated at an accelerating voltage of 200 kV. The data were recorded by a 2048×2048 CCD camera (Gatan). The defocus was ~2 mm. The software Bsoft was used for standard image processing (164). Individual particles (AaLS-13: 636 particles, AaLS-13 with HIV protease-R10: 854 particles) were picked using the software x3d (165). Averaged images were generated by reference-based alignments using the software SPIDER (166). For both samples, alignments were performed using a single reference image and four iterative rounds of averaging were carried out until the averaged images did not change anymore. For AaLS-13 coproduced with HIV protease-R10, additional alignments were performed using two reference images with representative “light” and “dark” interiors. This process groups the particles according to their greater similarity to one reference or the other, aligns and then averages them. The images were subjected to five iterative rounds of this sorting, aligning, and averaging until the averaged images did not change. The averaged images were normalized by standardizing the background density and the peak density of the capsid shell. Gray values (signal intensity) were (was) measured along eight equally spaced cross-sections of each averaged image using the program ImageJ. These cross-sections were then averaged and plotted as a function of distance from the capsid center (which corresponds to zero).
7.11 Mass spectrometry

Protein production was carried out as described in the “Protein production and purification” section. AaLS-capsids were purified first by Ni$^{2+}$-NTA affinity chromatography. The 500 mM imidazole fractions were then dialyzed against buffer A (50 mM sodium phosphate, 5 mM EDTA, pH 8.0) and further purified by anion exchange chromatography (Mono Q HR 10/10, Pharmacia) with a 160 ml linear gradient of buffer A to buffer B (50 mM sodium phosphate, 5 mM EDTA, 1 M NaCl, pH 8.0). The capsid-containing fractions were pooled, concentrated using a Microcon YM-50 filter (Millipore), and subjected to size exclusion chromatography on a Superose 6 10/300 GL column (GE Healthcare) eluted with buffer containing 50 mM sodium phosphate, 5 mM EDTA, 200 mM NaCl, pH 8.0. The fractions containing capsid were pooled and concentrated to 50-100 µl using a Microcon YM-100 filter (Millipore). Final protein concentrations were in the range 2.9 to 4.0 mg/ml.

For mass spectrometric analysis, 50 µg of protein was incubated in a 50 µl reaction in 0.1% RapiGest (Waters AG) for 10 min at RT, followed by addition of DTT to a final concentration of 10 mM and incubation for 30 min at 60°C. Samples were cooled on ice for 5 min and iodoacetamide was added to a final concentration of 25 mM. After incubation for 30 min at RT in the dark additional DTT was added to a final concentration of 35 mM. The trypsin digest was performed by adding trypsin in a ratio of 1:20 (w/w) and incubation of the sample for 3 h at 37°C. The reaction was quenched by addition of 10 µl 5% TFA in 50% acetonitrile and incubation for an additional 30 min. After centrifugation for 10 min at 13,000 g, the peptide-containing supernatant was subjected to ZipTip (Millipore) purification according to the manufacturer’s instructions. After vacuum concentration samples were redissolved in 15 µl 3% acetonitrile containing 0.1% formic acid. LC-MS/MS analysis was performed on an LTQ-Orbitrap XL (ThermoFisher Scientific). Samples were injected into an Eksigent-nano-HPLC system (Eksigent Technologies) by an autosampler and separated on a self-made reverse-phase tip column (75 µm × 80 mm) packed with C18 material (3 µm, 200 Å, AQ, Bischoff GmbH). The column was equilibrated with 97% solvent A (A: 1% acetonitrile; 0.2% formic acid in water)
and 3% solvent B (B: 80% acetonitrile, 0.2% formic acid in water). Peptides were eluted using the following gradient: 0-5 min, 3-10% B; 5-55 min, 10-40% B; 55-60 min, 40-97% B at a flow rate of 0.2 μl/min. High accuracy mass spectra were acquired in mass range of 300-2000 m/z and a target value of 5x10^5 ions and a resolution of 60,000 at 400 m/z. Up to five data-dependent MS/MS were recorded in parallel in the linear ion trap of the most intense ions with charge state 2+ or 3+ using collision induced dissociation. Target ions already selected for MS/MS were dynamically excluded for 60 s.

After data collection the peak lists were generated using Mascot Distiller software 2.3.12 (Matrix Science Ltd.) and searched against the Swissprot database extended with the specific protein entries for AaLS-13, AaLS-wt, and the HIV protease-R10 (LINK) using the Mascot search algorithm (Mascot 2.3, Matrix Science Ltd.). Mascot was searched with a fragment ion mass tolerance of 0.6 Da and a parent ion tolerance of 7.0 ppm. The iodoacetamide derivative of cysteine was specified as a fixed modification, whereas oxidation of methionine, acetylation of protein N-termini, and pyro-glu formation of N-terminal glutamine residues were specified in Mascot as variable modifications. Scaffold (version Scaffold_3_00_04, Proteome Software Inc.) was used to validate MS/MS-based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 95.0% probability as specified by the Peptide Prophet algorithm (167). Protein identifications were accepted if they could be established at greater than 99.0% probability and contained at least 3 identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm (168). Relative quantification was done according to Grossmann et al. (98) using the Top 3 Protein Quantification (T3PQ) but all survey mass spectra were deconvoluted and for each protein the average of the three identified peptides with the largest peak areas was used to quantify the protein.
7.12 Circular dichroism (CD) spectroscopy

The secondary structure content of AaLS-neg and AaLS-13 were compared using far-UV CD spectroscopy. CD spectra of AaLS variants (10 μM) were recorded at 25 °C in 50 mM sodium phosphate buffer, 200 mM NaCl, pH 8, in a cuvette with a pathlength of 0.2 cm using an Aviv 202 CD Spectrometer (Aviv Biomedical). Spectra were obtained by averaging 3 wavelength scans from 200 to 260 nm in 1 nm steps, with a signal averaging time of 2 s and a bandwidth of 1 nm.

Thermal denaturation curves were determined using again a protein concentration of 10 μM and a 0.2 cm cuvette. The ellipticity at 222 nm was monitored as a function of temperature from 25 °C to 99 °C in 1 °C steps with an averaging time of 2 s and a rise of 3 °C per minute.

7.13 Encapsulation efficiencies for GFP variants

GFP fluorescence of Ni²⁺-affinity purified AaLS capsids coproduced with the GFP variants was measured on a Photon Technology International fluorimeter. The samples were adjusted to a total protein concentration of 0.05 mg/ml in 50 mM sodium phosphate buffer, 300 mM NaCl, pH 8 and GFP was excited at 450 nm and the emission was measured from 470 to 600 nm with 1 nm step size, and a bandpass of 4 nm for excitation and 1 nm for emission. The value at the emission wavelength of maximal fluorescence was used. To determine the fluorescence efficiency of GFPpos compared to GFP-R₁₀ the fluorescence of free GFPpos and GFP-R₁₀ of known concentrations, determined using the Coomassie Plus Protein Assay Reagent (Pierce) with bovine serum albumin (BSA) as the standard, were measured. GFPpos has a 1.75 fold lower fluorescence than equal amounts of GFP-R₁₀. Fluorescence of samples with GFPpos were corrected by this factor of 1.75.
7.14 Dissociation of AaLS-guest complexes

200 µl aliquots of Ni\textsuperscript{2+}-affinity purified AaLS capsids (1-2.5 mg/ml) coproduced with HIV protease-R\textsubscript{10} or the GFP variants were injected at different time intervals on a size exclusion Superose 6 10/300 GL column (GE Healthcare) run with buffer containing 50 mM sodium phosphate, 5 mM EDTA, 200 mM NaCl, pH 8.0. 0.5 ml fractions between an elution volume of 5 and 25 ml were collected.

HIV protease in these fractions was detected by ELISA. Aliquots of the fractions were denatured in SDS gel loading buffer and then 10-fold dilutions in carbonate buffer coated onto Immunomaxi 96 well plates (TPP). HIV protease was detected as describes in the “ELISA measurement” section. The GFP fluorescence of 250 µl of the size exclusion fractions was measured in 96 well plates (96F Nuclon Delta Black Microwell SL, Nunc) on a fluorescent plate reader (Varioskan, Thermo Fischer) with exitation wavelength of 480 nm, and emission at 507 nm and a measurement time of 1000 ms.

7.15 In vitro loading

AaLS-13 capsids were purified by Ni\textsuperscript{2+}-affinity chromatography and 5 mM EDTA was immediately added to the eluted proteins in some samples, but omitted from others for comparison. Samples were subjected to size exclusion chromatography on Sephacryl S-300 column (Amersham Pharmacia) eluted with buffer containing 50 mM sodium phosphate, 200 mM NaCl, pH 8.0. Fractions corresponding to quaternary states smaller than capsids and to intact capsids were separately pooled and concentrated using a Microcon YM-30 or YM-100 filter (Millipore), respectively. Protein concentrations were determined using Coomassie Plus Protein Assay Reagent (Pierce). For in vitro loading the AaLS-13 samples were mixed with GFPpos in different molar ratios, with final concentrations of AaLS-13 monomers in the range 85 to 170 µM, and EDTA at a final concentration of 5 mM at a reaction volume of 600 µl.
The reassembly and loading process was monitored by periodically injecting aliquots of the loading mixtures on a Superose 6 10/300 GL (GE Healthcare) size exclusion column at different time intervals run with buffer containing 50 mM sodium phosphate, 5 mM EDTA, 200 mM NaCl, pH 8.0. GFP was detected in the size exclusion fractions as described in the “Dissociation of AaLS-guest complexes” section.

7.16 Kinetic measurements for retro-aldol cleavage of (±)-methodol

Fluorescence measurements for initial investigation of the retro-aldol activity of the ScLS designs were conducted on a Photon Technology International fluorimeter. Retro-aldol cleavage of (±)-methodol ((±)-(4-hydroxy-4-(6-methoxy-2-naphthyl)-2-butanone) was monitored by measuring fluorescence of the product 6-methoxy-2-naphthylaldehyde from 400 to 500 nm with 1 nm step size at an excitation wavelength of 330 nm, and 500 s waiting time between measurement of each spectrum. The background reaction was measured with 3500 s waiting time between spectra for 16 h. The reactions were performed with 500 µM (±)-methodol in 50 mM sodium phosphate, pH 7.5 containing 150 mM NaCl and 5 % MeCN at 25 °C and a protein concentration of 20 µM ScLS-pentamer. Initial velocities were determined by linear regression of the emission at 452 nm. Reaction rates were corrected for background retro-aldol cleavage. Substrate conversions were calculated from the initial velocities using a calibration curve of 6-methoxy-2-naphthylaldehyde.

For Michaelis-Menten kinetics and pH rate profiles fluorescence measurements were carried out on a Fluoroskan Ascent FL (Thermo Scientific) in 96 well plates (96F Nuclon Delta Black Microwell SI, Nunc). Product formation was recorded at 450 nm and an excitation wavelength of 335 nm every 2 min. Michaelis-Menten kinetics were measured with 125-2000 µM (±)-methodol. Data were fit to the Michaelis-Menten equation \( \frac{v_0}{[E]} = \frac{k_{cat}[S]}{(K_M + [S])} \). pH rate profiles were measured with 100 µM (±)-methodol using the buffers: 50 mM citrate at pH 5, 50
mM phosphate from pH 6 to 8, and 50 mM 2-amino-2-methyl-1-propanol-HCl from pH 9 to 10.

7.17 96 well plate screening assay for retro-aldol cleavage of (±)-methodol

Single colonies of BL21 or BL21-gold cells that had been transformed with plasmids containing the ScLS variants (pMG-ScLS-wt, pMG-ScLS-2), RA45.2 (pET-29b(+) -RA45.2) or no plasmid were used to inoculate precultures in 96-well microtiter plates (TC Microwell 96U w/lid Nuclon D SI, Nunc) containing 150 µl LB supplemented with 150 µg/ml ampicillin (ScLS variants) or 30 µg/ml kanamycin (RA45.2) per well. The plates were covered with air permeable membranes (Breathe Easy, Diversified Biotech) and incubated overnight at 25 °C and 270 rpm.

96-deep-well plates (Deepwell plate 96/2000µl, Eppendorf), filled with 2 ml LB supplemented with the appropriate antibiotic, were inoculated with 20 µl preculture, sealed with air permeable membranes and incubated at 30 °C and 250 rpm. After ca. 10 h, when the OD600 had reached 0.5-0.7, protein production was induced with IPTG (0.1 mM). 20 h after induction, the cells were harvested by centrifugation (10 min, 4000 rpm, 4 °C) and the supernatant was removed by inverting the plates. Some of the cell pellets were washed one time with 1 ml buffer (50 mM sodium phosphate, 150 mM NaCl, pH 7.5). The pellets were stored overnight at -20 °C.

The pellets were resuspended in 300 µl buffer (50 mM sodium phosphate, 150 mM NaCl, pH 7.5) containing 1 mg/ml lysozyme and incubated for 1 h at room temperature. The plates were shaken 4 °C and 800 rpm for 5 minutes and then subjected to five freeze-thaw cycles involving freezing at -80 °C for 30 min, thawing for 15 min at 23 °C, and shaking for 15 min at 800 rpm at 4 °C. After centrifugation for 20 min at 4 °C and 4000 rpm, 190 µl of the clear supernatants were transferred into 96 well plates (96F Nuclon Delta Black Microwell SI, Nunc). 10 µl (±)-methodol in MeCN was added to give a final substrate concentration of 500 µM. Using a fluorescence plate reader (Varioskan, Thermo
Fischer) retro-aldol cleavage of (±)-methodol was monitored at 452 nm and 25 °C, with an excitation wavelength of 330 nm.
8 REFERENCES


 Structural insights and potential as a protein expression tool. M. Furutani


82. W. Meining et al., The atomic structure of pentameric lumazine synthase from *Saccharomyces cerevisiae* at 1.85 angstrom resolution reveals the binding mode of a phosphonate intermediate analogue. *J. Mol. Biol.* **299**, 181 (2000).
98. J. Grossmann et al., Implementation and evaluation of relative and absolute quantification in shotgun proteomics with label-free methods. J. Proteomics 73, 1740 (2010).


118. P. Kast et al., A strategically positioned cation is crucial for efficient catalysis by chorismate mutase. J. Biol. Chem. 275, 36832 (2000).


CURRICULUM VITAE

BIGNA WÖRSDÖRFER

Laboratory of Organic Chemistry, ETH Zürich, Höngetterberg HCI F336, 8093 Zürich, Switzerland
Phone: +41 44 63 34201
Email: woersdoerfer@org.chem.ethz.ch

Nationality: German
Date of birth: May 22nd, 1980

Education

Dec. 2004-Present  PhD thesis studies in Chemistry
ETH Zurich, Switzerland

2004  Diploma in Biochemistry
University of Bayreuth, Germany

1999  Abitur (general qualification for university entrance)
Freiherr-vom-Stein-Gymnasium in Fulda, Germany

Research experience

2004-Present  PhD thesis
“Functionalizing capsid forming proteins”
Laboratory of Prof. D. Hilvert, Laboratory of organic chemistry, ETH Zurich, Switzerland

2003-2004  Diploma thesis
“Expression of glycosylated fusionproteins in Leishmania terentolae.”
Laboratory of Prof. C. Unverzagt, Laboratory of Bioorganic Chemistry, University of Bayreuth, Germany

2003  Undergraduate research (abroad, 1 ½ months)
“Colocalisation studies of the actin-bundling protein fascin and b-catenin or E-cadherin in mammalian cells by immunofluorescence microscopy.”
Laboratory of Prof. A. Yap, Molecular Cell Biology, Institute for Molecular Bioscience, Brisbane, Australia