

Author Manuscript

Title: The Structural Diversity of Encapsulin Protein Shells

Authors: Tobias W Giessen, Ph.D.

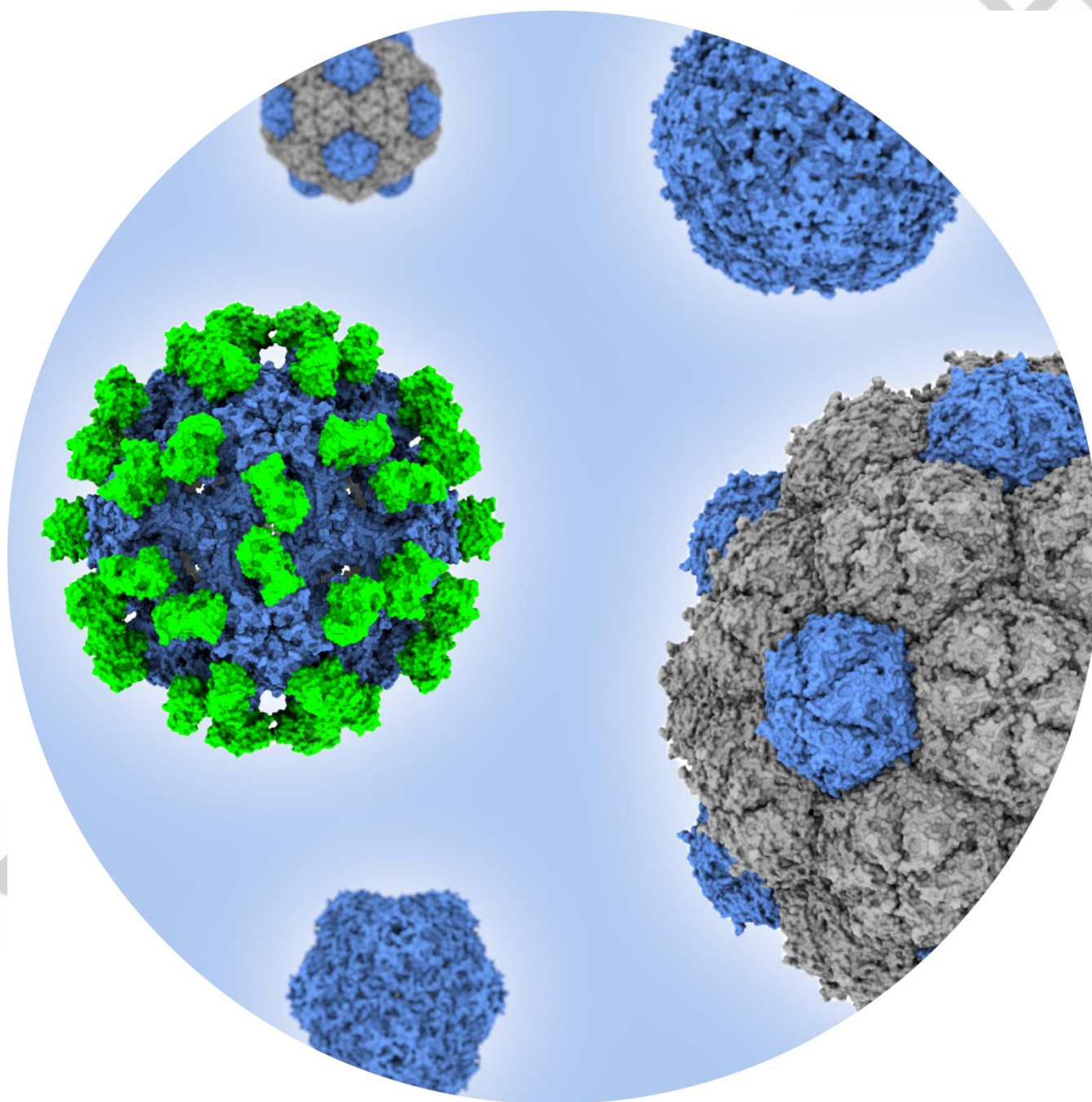
This is the author manuscript accepted for publication. It has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record.

To be cited as: 10.1002/cbic.202400535

Link to VoR: <https://doi.org/10.1002/cbic.202400535>

The Structural Diversity of Encapsulin Protein Shells

Tobias W. Giessen*^[a]



Author Manuscript

[a] T. W. Giessen
 Department of Biological Chemistry
 University of Michigan, Ann Arbor
 1150 W Medical Center Dr, Ann Arbor, MI 48109-5622
 E-mail: tgiessen@umich.edu

Abstract: Subcellular compartmentalization is a universal feature of all cells. Spatially distinct compartments, be they lipid- or protein-based, enable cells to optimize local reaction environments, store nutrients, and sequester toxic processes. Prokaryotes generally lack intracellular membrane systems and usually rely on protein-based compartments and organelles to regulate and optimize their metabolism. Encapsulins are one of the most diverse and widespread classes of prokaryotic protein compartments. They self-assemble into icosahedral protein shells and are able to specifically internalize dedicated cargo enzymes. This review discusses the structural diversity of encapsulin protein shells, focusing on shell assembly, symmetry, and dynamics. The properties and functions of pores found within encapsulin shells will also be discussed. In addition, fusion and insertion domains embedded within encapsulin shell protomers will be highlighted. Finally, future research directions for basic encapsulin biology, with a focus on the structural understanding of encapsulins, are briefly outlined.

1. Introduction

Intracellular compartmentalization is one of the defining features of life.^[1, 2] As prokaryotes do generally not possess lipid-based organelles, they have evolved alternative, often protein-based, strategies to compartmentalize and spatially organize their cell interiors.^[3-5] One such approach, widely found across many bacteria and archaea, are protein compartments – self-assembling protein structures consisting of a protein shell and encapsulated enzymes or whole metabolic pathways.^[6] Physical sequestration of biochemical reactions facilitates their spatial and temporal regulation.^[7-9] Compartmentalization in general enables the dynamic storage of nutrients and the sequestration of unwanted compounds, the transport of molecules within or between cells, and the creation of distinct reaction spaces optimized for a given biological process.^[10, 11]

Bacterial microcompartments (BMCs) were the first class of protein compartments recognized in prokaryotes.^[5] BMC protein shells consist of multiple kinds of hexameric, pentameric, and trimeric shell protomers, assemble into (quasi)-icosahedral structures, and can exceed diameters of 100 nm.^[12, 13] BMCs either fulfill an anabolic role as carbon-fixing organelles called carboxysomes^[14, 15] or participate in catabolic pathways as metabolosomes involved in carbon and nitrogen source utilization.^[16-18]

More recently, another widespread protein compartment type, called encapsulin nanocompartments or encapsulins for short, has been discovered.^[8, 19, 20] Encapsulins self-assemble into generally icosahedral protein compartments with diameters between ca. 20 and 50 nm.^[20-22] The eponymous feature of encapsulins is their ability to encapsulate dedicated cargo

enzymes.^[23-26] Two different cargo loading mechanisms have been identified so far. First, native cargo proteins can contain short and conserved, ten to twenty residues long, peptide sequences at their termini – called targeting peptides (TPs) or cargo-loading peptides (CLPs) – that mediate cargo loading during shell self-assembly.^[23] Alternatively, cargo proteins can contain extended, mostly disordered domains, often hundreds of residues in length, exhibiting short, five to ten residues long, conserved motifs that facilitate cargo encapsulation.^[27-29] Encapsulins are found in many bacterial and archaeal phyla and are involved in oxidative stress resistance,^[25, 30-32] iron metabolism and storage,^[20-22, 33-35] sulfur metabolism,^[27, 29] and secondary metabolite biosynthesis.^[28, 36] What differentiates encapsulins from BMCs and other well-known protein cages like ferritin or lumazine synthase is the fact that the encapsulin shell protein shares the HK97 phage-like fold, suggesting an evolutionary connection between encapsulins and viruses.^[8, 19, 20]

This review focuses on the structural diversity of encapsulin protein shells. Encapsulin biology,^[8, 9, 37] the structural and mechanistic details of encapsulin cargo loading,^[23] and advances in encapsulin engineering^[38-46] have recently been reviewed and will not be covered here. Special focus will be placed on encapsulin shell symmetry, composition, and dynamics. **Shell pores, which mediate the selective movement of small molecule substrates and products of encapsulated cargo enzymes into and out of the shell, will also be discussed. Finally, fusion and insertion domains found within the encapsulin shell protomer, whose functions largely remain to be elucidated, will be highlighted. In closing, future research directions will be briefly outlined.**

Tobias W. Giessen obtained his Diplom (BS/MS) in Chemistry at Phillips University Marburg in 2010 and his PhD in Biochemistry at the same institution in 2013, working with Mohamed A. Marahiel. Following postdoctoral research at Harvard Medical School and the Wyss Institute at Harvard in the group of Pamela A. Silver, Tobias started his independent career at the University of Michigan, Ann Arbor in 2019. His research interests are currently focused on encapsulin biology and engineering, enzyme filaments, and natural product biosynthesis.



2. Encapsulin Protomer and Shell Structure

Based on sequence and structural similarity as well as operon organization and genome neighborhood conservation,

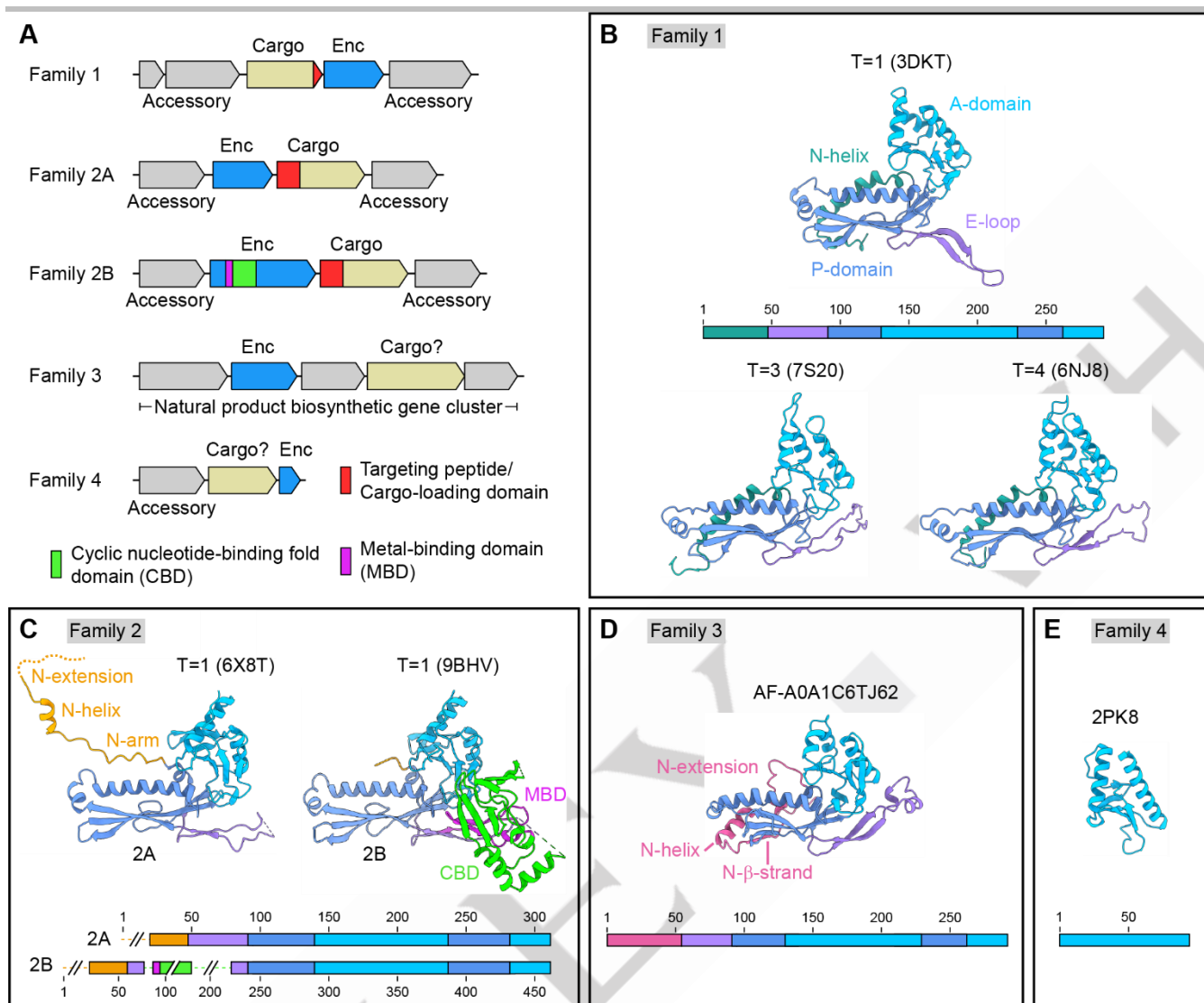


Figure 1. Encapsulin classification and protomer structures. (A) Encapsulin classification scheme based on sequence and structural homology as well as genome neighborhood conservation. Accessory genes represent non-cargo but conserved operon components. Question marks indicate lack of experimental evidence. Enc: encapsulin shell protein. (B) Structures of Family 1 encapsulin protomers. Examples of protomers found in T=1, T=3, and T=4 shells are shown. Core domains of the HK97 phage-like fold (A-domain, P-domain, and E-loop) are colored consistently throughout all panels. PDB IDs for all structures shown in parentheses. (C) Family 2 encapsulin protomers. A Family 2A protomer (left) and a Family 2B protomer (right) are shown. Distinct domains are highlighted by color. (D) Alpha Fold (AF)-predicted structure of a putative Family 3 encapsulin shell protein. A unique N-terminal extension consisting of an N-helix and an N-β-strand are highlighted in pink. UniProt ID shown above the protomer. (E) Family 4 encapsulin protomer structure.

encapsulins have been grouped into four families (Figure 1A).^[8, 19] Family 1 contains the best studied encapsulins, with experimental data available for multiple systems regarding their shell structures, associated cargo function, and cargo loading mechanisms. Family 1 encapsulins are present in over 30 prokaryotic phyla, with the majority found in Pseudomonadota, Actinomycetota, and Bacillota. Cargo loading in Family 1 systems is almost always based on short (ca. 10 residues) C-terminal targeting peptides. Recently, the first experimental data focused on Family 2 encapsulins have begun to emerge.^[27-29, 36] Family 2 can be divided into two sub-families based on the absence (Family 2A) or presence (Family 2B) of a cyclic nucleotide-binding fold domain (CBD)^[47] and a recently identified metal-binding sub-domain (MBD),^[28] both inserted into the HK97 phage-like fold. However, in comparison to Family 1 systems, little information is available for Family 2 encapsulins. Similar to Family 1 systems,

Family 2 encapsulins are broadly found across prokaryotic phyla and are especially common in members of the Actinomycetota and Pseudomonadota. Cargo loading in Family 2 encapsulins is mediated by extended (ca. 50-300 residues) N- or C-terminal cargo loading domains. Family 3 encapsulin systems currently lack experimental validation and remain putative. Family 3 encapsulins are always embedded in secondary metabolite biosynthetic gene clusters, primarily found in members of the Actinomycetota and Myxococcota. No obvious targeting peptides or domains have been identified in the surrounding operon components. Family 4 encapsulins represent heavily truncated versions of the HK97 phage-like fold and one Family 4 structure, determined as part of a structural genomics project, is currently available. Family 4 systems have been identified in the archaeal phylum Euryarchaeota and the bacterial phylum Bacteroidota.

REVIEW

Family 1 (Figure 1B), Family 2 (Figure 1C), and Family 3 (Figure 1D) encapsulin shell protomers exhibit strong structural homology to HK97-type viral capsid proteins^[8, 19] and share three universally conserved structural features – the axial domain (A-domain), peripheral domain (P-domain), and the extended loop (E-loop). In contrast, Family 4 protomers exhibit a heavily truncated HK97 phage-like fold, with only the triangular A-domain still being present (Figure 1E). All Family 1 encapsulin shell proteins contain protomers can contain an extended N-terminus consisting of an N-arm, N-helix, and an up to eighty residues long disordered N-extension (Figure 1C), features similar to the N-termini found in many HK97-type viral capsid proteins.^[29] While all Family 2 protomer sequences contain this extended N-terminus, in some Family 2B shell structures, it is not resolved (Figure 1C).^[28, 49] The abovementioned CBD and MBD insertions that differentiate 2B from 2A systems are both embedded into the E-loop with the MBD located below the CBD, making substantial contacts with the A-domain and P-domain. No structural data for Family 3 encapsulins is currently available. However, based on Alpha Fold predictions,^[50] the HK97 phage-like fold is conserved in Family 3 protomers (Figure 1D). An about fifty residues long N-terminal extension, dissimilar to Family 1 or Family 2 systems, is predicted to be present in Family 3 shell proteins. Parts of this extension form a fourth β -strand extending the three-strand P-domain β -sheet while a short helix occupies a similar position as the N-helix in Family 1 protomers. In Family 4 encapsulins, only the A-domain of the HK97 phage-like fold is present (Figure 1E). The single available Family 4 encapsulin structure suggests that Family 4 protomers form dimers and not protein shells. However, their structural homology, co-localization within operons with conserved enzyme components, and the fact that they are part of the same Pfam clan as the other encapsulin families (CL0373),^[51] has resulted in their annotation as a divergent class of encapsulins.

All studied Family 1 and Family 2 encapsulins assemble into closed protein shells, generally exhibiting icosahedral symmetry. These encapsulin shells share many characteristics with viral capsids assembled from HK97-type protomers.^[8] The HK97 phage-like fold is found in the major capsid proteins of bacteriophages of the order Caudovirales,^[52, 53] in capsid proteins of tailed archaeal viruses,^[54] and in the floor domains of animal-infecting viruses of the order Herpesviridae.^[55] Encapsulins represent the only non-viral assemblies known to exhibit the HK97 phage-like fold.^[8, 19, 20] The structural versatility and plasticity of the HK97 phage-like fold is highlighted by the fact that its three core structural features, the A-domain, P-domain, and E-loop, are conserved in proteins with low sequence similarity while a range of insertion and extension domains can be embedded at various positions within HK97-type proteins.^[56, 57] So far, Family 1 encapsulin shells with T=1, T=3, and T=4 icosahedral symmetry have been characterized consisting of 60, 180, or 240 protomers, respectively (Figure 2A). All icosahedral protein shells or capsids contain twelve five-fold, twenty three-fold, and thirty two-fold axes of symmetry. While T=1 shells can be assembled from twelve pentamers alone, icosahedra with T numbers greater than T=1 require both pentamers and hexamers to tile a closed shell or capsid.^[58, 59] In addition to twelve pentamers, T=3 shells contain twenty, and T=4 shells thirty, hexamers. T numbers also reflect the number of unique protomers occupying quasi-equivalent positions within the asymmetric unit (ASU) of a given protein shell, one for T=1, three for T=3, and four for T=4 shells (Figure 2B).

an N-terminal helix (N-helix) that together with the P-domain, forms the TP binding site, crucial for directing and anchoring cargo proteins to the interior of the shell.^[4, 24-26, 48] The primary structural difference between Family 1 protomers is the conformation of the E-loop, which can either point away from the A-domain in encapsulin shells with a triangulation number (T number) of T=1, or towards the A-domain in T=3 and T=4 shells (Figure 1B).^[20, 22, 33] In contrast to Family 1 systems, Family 2 The structures of distinct protomers within T=3 and T=4 ASUs are highly similar. This quasi-equivalency, first introduced by Caspar and Klug,^[58] as well as subtle shifts between ASUs, are responsible for a single type of shell protomer being able to form both pentameric and hexameric interfaces resulting in shells with different T numbers. While pentamers always align with the icosahedral five-fold axes of symmetry, hexamers can either be arranged around the three-fold (T=3) or two-fold (T=4) symmetry axes (Figure 2C). All so far characterized Family 2 encapsulins, both 2A and 2B systems, exhibit T=1 icosahedral symmetry (Figure 2D).^[27-29, 49] In comparison to Family 1 shells, Family 2A encapsulins have a less spherical appearance caused by a more pronounced turret-like conformation of the twelve shell pentamers. Family 2B shells exhibit a similar assembly of the core HK97 phage-like fold as compared to Family 2A systems, however, the presence of the CBD and MBD insertions results in a strikingly different appearance. CBD/MBD domains are externally displayed, and one pair of CBD/MBD domains is arranged around all thirty two-fold axes of symmetry (Figure 2D).^[28, 49]

While most Family 1 encapsulin shells are currently assumed to exclusively function as selectively permeable diffusion barriers, the *Thermotoga maritima* T=1 shell has been found to contain a flavin, likely flavin mononucleotide (FMN), binding site (Figure 2E).^[60, 61] FMNs are bound at protomer interfaces close to the three-fold axis of symmetry for a total of sixty flavins per 60mer T=1 shell. The FMN binding site is formed by a conserved tryptophan residue and an arginine residue sandwiching the FMN isoalloxazine ring via π -stacking interactions (Figure 2E). The precise function of this FMN in the context of the iron-precipitating *T. maritima* encapsulin is currently unknown. However, it has been proposed to play a role in electron transport through the shell to enable the reduction and release of stored iron.^[61, 62]

Nearly all characterized HK97 phage-like fold assemblies are formed by a single major capsid or shell protein. There is one characterized virus – bacteriophage T4 – that utilizes two distinct HK97-type capsid proteins for distinct structural roles, forming either homo-pentamers or homo-hexamers.^[63, 64] Genome-mining studies have identified many Family 2B encapsulin operons encoding two distinct encapsulin shell proteins.^[8, 19] Recently, it was shown that such a Family 2B system found in *Streptomyces lydicus* does indeed form two-component mixed shell assemblies with variable shell composition (Figure 2F).^[49] In contrast to the bacteriophage T4 capsid, the two *S. lydicus* encapsulin shell components form a quasi-icosahedral T=1 shell where both shell protomers can occupy any position and engage in all possible inter-protomer interactions. The ability of these two protomers to form hetero-pentamers with variable relative compositions results in the irregular tiling of the T=1 shell, thus deviating from strict icosahedral symmetry.

As mentioned above, Family 4 encapsulins do not seem to form icosahedral shells but instead dimeric assemblies (Figure 2G). The interaction between the two Family 4 proteins is different from the interaction between two adjacent A-domains found in the

REVIEW

context of an HK97-type shell or capsid. This is not surprising as most features of the HK97 phage-like fold have been lost in Family 4 encapsulins.

3. Pores in Encapsulin Shells

All so far studied Family 1 and Family 2 encapsulin shells contain pores. As is the case for other protein compartments like ferritins and BMCs, encapsulin pores are thought to be crucial for controlling the flux of small molecules and ions into and out of the encapsulin shell. Encapsulin pores are generally smaller than 20 Å in diameter which generally prevents any protein from gaining access to the encapsulin lumen once the shell is assembled (Figure 3A).^[20-22, 25, 26, 32, 33, 48, 60, 61, 65-67] Static pores smaller than ca. 4.5 Å are essentially closed for small molecule flux but may

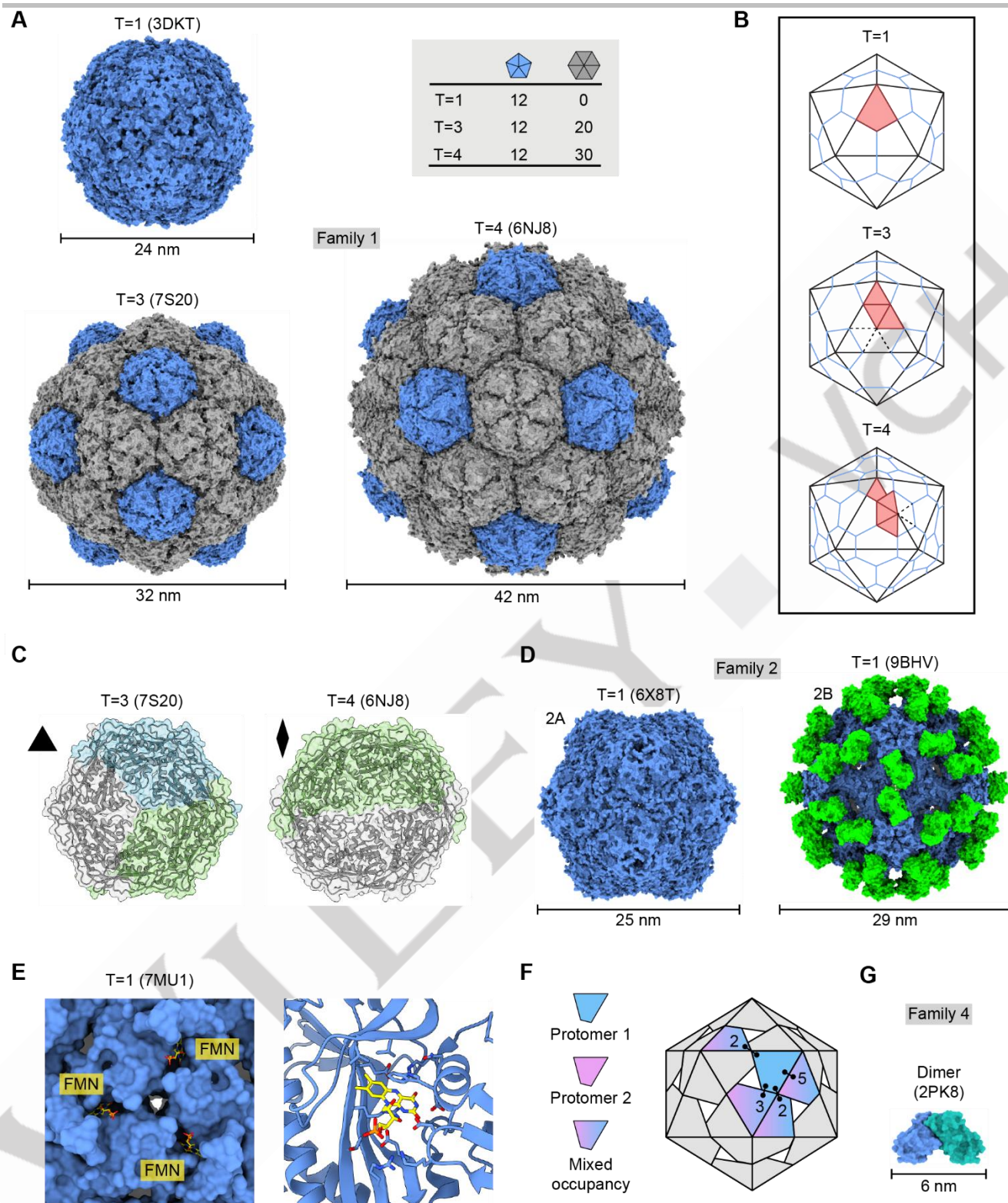


Figure 2. The diversity of encapsulin assemblies. (A) Family 1 encapsulin shells. T=1, T=3, and T=4 shells are shown. Pentamers are colored in blue, hexamers in grey. The number of pentamers and hexamers necessary to form a shell of a given triangulation (T) number is shown. (B) The symmetry and organization of icosahedral encapsulin shells with different T numbers. Asymmetric units (ASUs) are shown in red. (C) T=3 (left) and T=4 (right) Family 1 hexamers are shown highlighting their ability to occupy both three-fold (T=3) and two-fold (T=4) symmetrical positions. (D) Family 2A (left) and 2B (right) shells. External CBD/MBDs are shown in green. (E) Family 1 encapsulin shell with a bound flavin mononucleotide (FMN) cofactor. FMN binding sites are arranged around the three-fold symmetry axis (left). FMN binding site (right). (F) Schematic of the complex assembly of mixed two-component encapsulin shells. Numbers highlight unique interactions around the 2-, 3-, and 5-fold axes (G) Structurally characterized dimeric assembly of a Family 4 encapsulin.

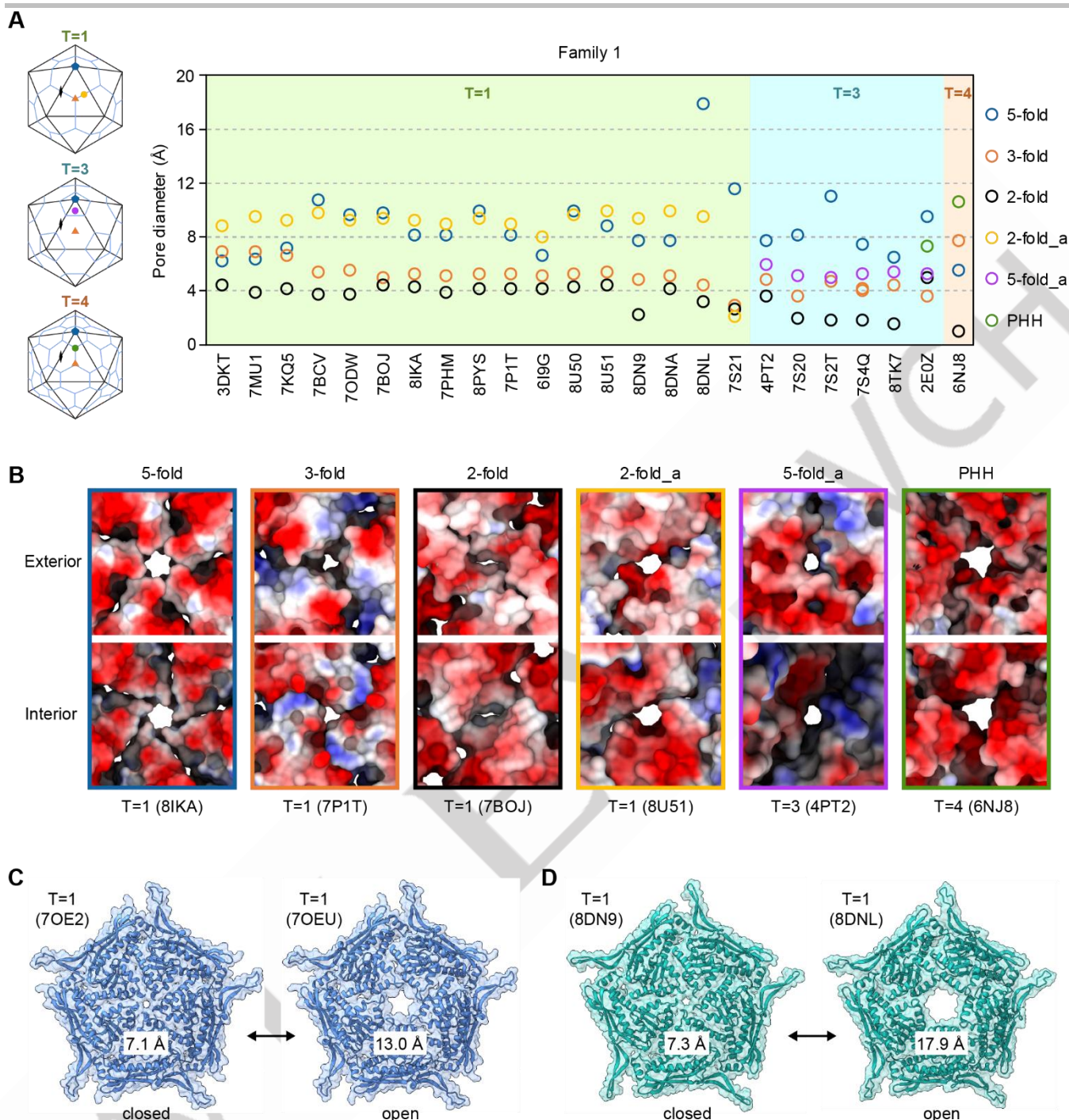


Figure 3. Pores in Family 1 encapsulins. (A) Identified pore positions in T=1, T=3, and T=4 encapsulin shells (left) and summary of all pores identified in structurally characterized Family 1 shells highlighting their pore diameters (right). PDB IDs shown on the x-axis. 2-fold_a: two-fold adjacent; 5-fold_a: five-fold adjacent; PHH: pentamer-hexamer-hexamer interface. (B) Representative examples of exterior and interior surface views of all pore types identified in Family 1 encapsulins. Electrostatic surfaces are shown. (C) Five-fold pore dynamics with discrete closed (left) and open (right) states identified in the *Haliangium ochraceum* T=1 encapsulin. (D) Pore dynamics of the *Acidipropionibacterium acidipropionici* T=1 shell.

still transmit inorganic ions.^[7, 68] Thus, the encapsulin shell acts as a selectively permeable diffusion barrier, allowing small molecules and ions, necessary for encapsulin function, to enter the shell via its pores while excluding all other, potentially interfering, cellular components from the shell interior. This can create specific nano-environments inside the encapsulin shell, optimized for a given process. For example, a recently characterized Family 2A encapsulin shell was found to exclude

the ubiquitous cellular reducing agent glutathione, thus creating a more oxidizing interior environment, necessary for the biological function of the system.^[27]

While open pores (>4.5 Å) at the five-fold axis of symmetry (5-fold) can be found in all so far characterized Family 1 encapsulin shells, open pores located at the three-fold symmetry axis (3-fold) are only found in some T=1 and T=4 shells (Figure 3A). Two-fold pores (2-fold) are generally very small and likely only able to

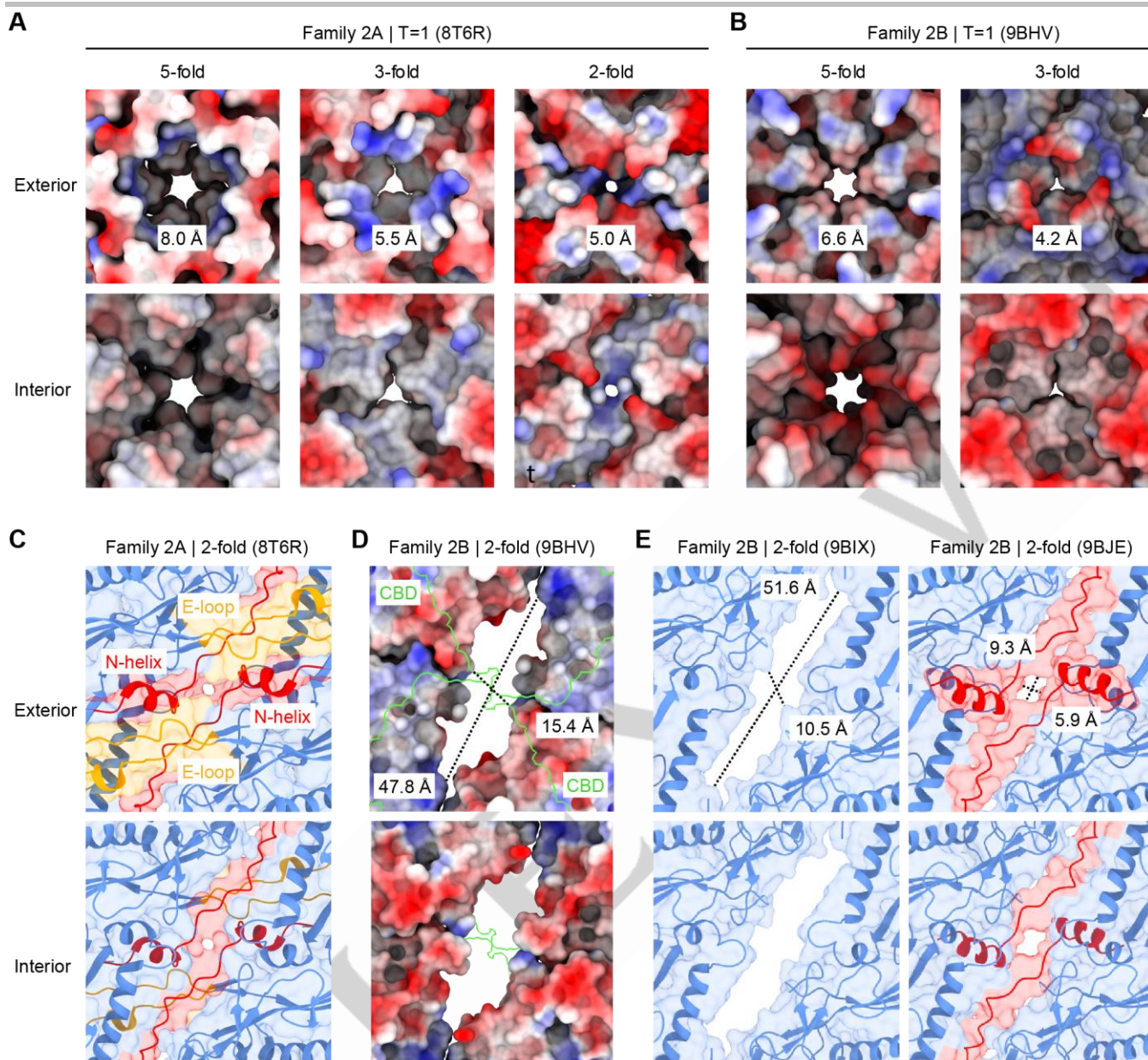


Figure 4. Pores in Family 2 encapsulins. (A) Exterior and interior surface views of all pore types identified in the *Acinetobacter baumannii* Family 2A encapsulin. Electrostatic surfaces are shown. (B) Five- and three-fold pores found in the *Streptomyces griseus* Family 2B shell are shown as electrostatic surface. (C) Detailed exterior and interior view of a two-fold pore found in a Family 2A shell. The N-arm and N-helix are shown in red. E-loops are shown in yellow. Both ribbon and surface representations are shown. (D) The open two-fold pore of the *S. griseus* Family 2B encapsulin. The position of CBDs above the two-fold pores is outlined in green. Besides their outline, CBDs are fully transparent. (E) The different two-fold pore states observed in the *Streptomyces lydicus* two-component Family 2B shell – open (left) and closed (right). CBDs omitted for clarity. The N-arm and N-helix are shown in red.

transmit ions. Another type of pore found exclusively in Family 1 T=1 shells are two-fold adjacent pores (2-fold_a). They often represent the largest pores in T=1 shells. A pore type only found in T=3 shells are five-fold adjacent pores (5-fold_a) located within pentamers. With T=3 and T=4 shells containing both pentamers and hexamers, they can form another pore type located at the pseudo-three-fold interface between a pentamer and two hexamers (PHH). Overall, encapsulin pore sizes range from essentially closed up to 18 Å in diameter. Across different Family 1 shell sizes and cargo types, the electrostatic charge at the exterior and interior of pores is mostly negative to neutral, rarely positive (Figure 3B). While the details of how pore size and charge

influence the functioning of native encapsulin systems remains to be elucidated, a number of mutational studies, focused on five-fold pore residues, have found that altering pore properties has a clear effect on the diffusion of small molecules and ions across the shell.^[7, 68-70] In addition, changing pore residues can often disrupt proper shell assembly or even result in alternative assembly geometries, deviating from icosahedral symmetry.^[71] This highlights the importance of pore residues for mediating shell assembly and overall shell stability. Recent structural analyses of Family 1 T=1 encapsulin shells have highlighted that five-fold pores are not static but dynamic. Discrete closed (7OE2) and open (7OEU) conformations could be

resolved for the *Haliangium ochraceum* T=1 encapsulin shell under physiological conditions with a substantial difference in pore diameter between the two states of ca. 6 Å (Figure 3C).^[65] Similarly, the T=1 encapsulin shell of the acid-tolerant bacterium *Acidipropionibacterium acidipropionici* was found to display two resolvable five-fold pore states – closed and open – at physiological pH while under strongly acidic conditions, only a single closed state could be observed (Figure 3D).^[67] The difference in pore diameter is over 10 Å, with the *A. acidipropionici* open pore state representing the largest so far identified pore in Family 1 encapsulin shells. In both of these systems displaying dynamic five-fold pores, a conformational change of the pore-loop connecting the protomer helices $\alpha 6$ and $\alpha 7$ within the A-domain, in combination with a concerted backwards tilt of the pentameric protomers away from the five-fold symmetry axis, allows for the switch between closed and open five-fold pore states.

As highlighted above, the HK97-type domain of Family 2A and 2B encapsulins is highly similar.^[27-29, 49] This similarity extends to their five- and three-fold pores (Figure 4A and 4B) which, at 4 to 8 Å in diameter, are comparable to Family 1 pores. In general, pores in Family 2 systems are less strongly negatively charged and the pore transits are relatively electrostatically neutral. A major difference between both Family 1 and Family 2 encapsulins as well as 2A and 2B systems are their two-fold pores. While the two-fold pores in Family 1 encapsulins are formed by the E-loops of two protomers, in Family 2 systems, two-fold pore structure depends on the N-arms of two neighboring shell proteins. In Family 2A encapsulins, the N-arms and N-helices of two protomers interact and surround the two-fold axis of symmetry, forming a relatively small two-fold pore (<4.5 Å) (Figure 4C).^[27, 29] The two N-arms/N-helices are stabilized by overlapping E-loops of neighboring protomers in a chainmail-like fashion.^[53] In contrast, in the *Streptomyces griseus* Family 2B shell (9BHV), a markedly different two-fold pore can be observed.^[28] The fact that the N-arm, N-helix, and N-extension are not structured in the *S. griseus* system results in a very large and elongated two-fold pore, with dimensions of ca. 15 × 50 Å (Figure 4D). While the externally displayed CBDs, inserted into E-loops, partially occlude this Family 2B two-fold pore when viewed from the shell exterior (Figure 2D), the pore is not affected by CBDs and is in fact completely open. In the recently characterized two-component Family 2B system from *S. lydicus*,^[49] two conformationally distinct two-fold pores states are present (Figure 4E), further adding to the complexity of this mixed quasi-icosahedral shell assembly. Specifically, an open state with unstructured N-terminus, similar to the Family 2B *S. griseus* system, and a closed state with clearly resolved N-terminus, reminiscent of Family 2A shells, could be observed.

4. Domain Fusions and Insertions

Beyond the ability to form icosahedral capsids and shells of vastly different size, ranging from T=1 to T=52, the HK97 phage-like fold is also able to accommodate a diverse range of different insertion and fusion domains.^[56, 72] For example, the bacteriophage T4 HK97-type capsid protein exhibits an externally displayed chitin-binding-like insertion domain within its E-loop^[73] while the bacteriophage 029 capsid protein contains an immunoglobulin-like fusion domain at its C-terminus.^[74] Bacteriophages P22,^[75, 76] Sf6,^[77] and CUS-3^[78] all contain external insertion domains within

the A-domains of their HK97 phage-like fold coat proteins. The most striking example of a C-terminal fusion, partially interwoven with the HK97 phage-like fold A-domain, can be found in viruses of the order Herpesviridae where large middle and upper domains protrude from the HK97-type floor domain to form over 10 nm high external towers.^[79-83] N-terminal fusions are also observed in HK97-type viral capsids – called Δ -domains – that act as fusion scaffolding domains enabling the formation of capsids with T number larger than 1.^[84, 85]

Similar to HK97-type viral capsid proteins, encapsulin shell proteins have also been found to contain a number of different types of insertion or fusion domains.^[8, 19, 21, 86] So far, only one type of domain – the CBD/MBD insertion found in Family 2B shells – has been structurally characterized in the context of the encapsulin shell.^[27, 49] As mentioned above, CBD/MBD insertions are embedded into the E-loop of the HK97 phage-like fold of the encapsulin shell protein at the genetic level. They are displayed externally and arranged around all two-fold symmetry axes of the icosahedral Family 2B shell (Figure 5A). The MBD can be described as a sub-domain within the CBD (Figure 5B) and makes substantial contacts with the A-domain of the HK97 phage-like fold. The MBD does not share any significant sequence or structural homology to any characterized domain. A metal ion – based on coordination geometry, likely a magnesium ion – is bound within the MBD by a set of four conserved residues located within a tight loop (Figure 5C). It has been hypothesized that the MBD may play a role in mediating CBD-HK97 phage-like domain interactions and might be important for controlling CBD conformation.^[27, 49] While CBDs share sequence and structural similarity with cyclic adenosine monophosphate (cAMP)-binding domains,^[87, 88] especially the catabolite activator protein (CAP) found in *Escherichia coli*,^[89, 90] Family 2B CBDs do not seem to bind cAMP.^[28] This is supported by multiple types of orthogonal *in vitro* binding experiments which all showed no detectable affinity of cAMP towards CBDs. In addition, analysis of the putative cAMP binding site indicates that a crucial and conserved arginine residue found in *E. coli* CAP (R82) and other confirmed cAMP-binding domains is missing in Family 2B CBDs (Figure 5D). Instead, a conserved tryptophan residue (W176) is found in the equivalent position, likely preventing cAMP binding. These facts indicate that Family 2B CBDs likely bind a, yet to be identified, small molecule ligand, different than cAMP. Regarding the function of CBDs, it has been suggested that ligand binding might result in a conformational change of the CBDs themselves and/or the E-loop, resulting in the opening or closing of the two-fold pores.^[28] Considering that distinct open and closed two-fold pore states have already been reported, albeit both in the absence of any detectable ligand, this hypothesis seems possible.^[49] By allowing or restricting substrate flux into the encapsulin shell, distinct pore states may modulate the activity of encapsulated cargo enzymes. In essence, the function of CBDs may be to receive an external signal, in the form of a small molecule ligand, which in turn is used by the Family 2B shell to control the activity of internalized cargo enzymes.

The very first encapsulin shell structure reported – at the time not recognized as an encapsulin but described as a virus-like particle – was a Family 1 T=3 encapsulin found in the hyperthermophilic archaeon *Pyrococcus furiosus* (2E0Z).^[21] This archaeal Family 1 encapsulin is now recognized to be part of a class of fusion encapsulins carrying N-terminal ferritin-like protein (Flp) fusion domains (Figure 6A).^[8, 19] As the Family 1 N-terminus is located

REVIEW

on the interior of the assembled shell, all N-terminal fusions will be localized to the encapsulin lumen, similar to non-fusion cargo proteins. No other separate cargo proteins have so far been

identified for archaeal Flp fusion encapsulins. In the original *P. furiosus* T=3 shell structure, determined via x-ray crystallography,

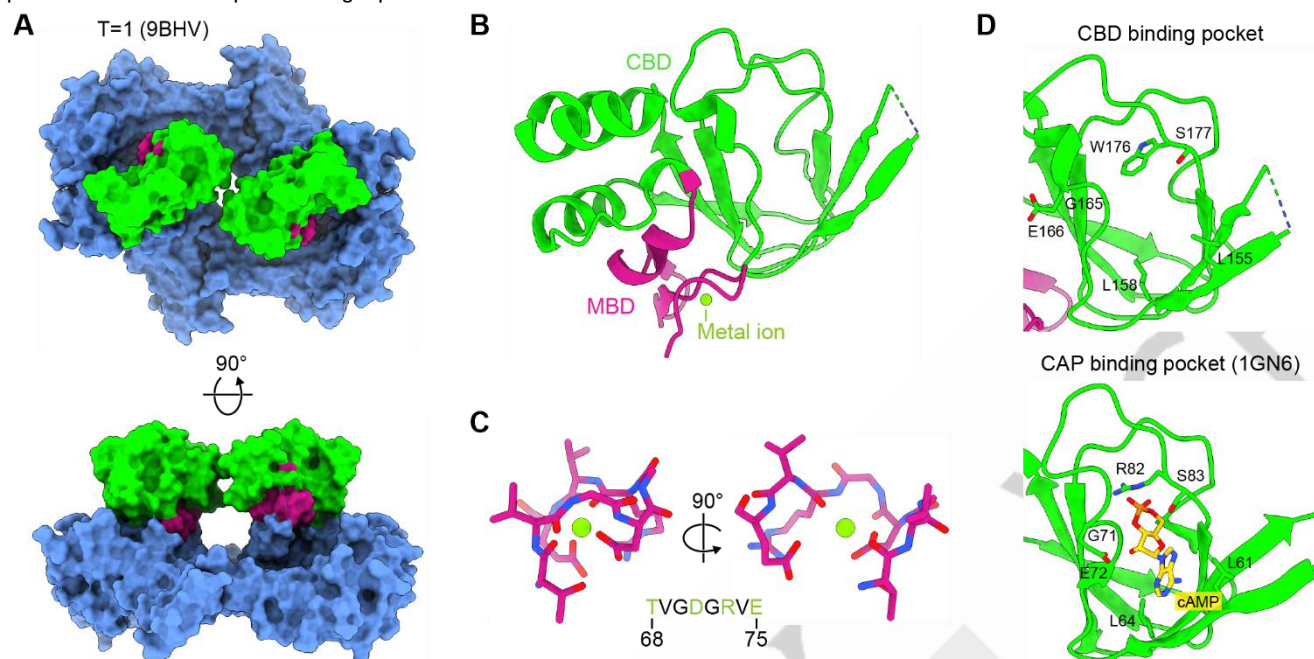


Figure 5. CBD and MBD insertions in Family 2B encapsulins. (A) Top and side views of externally displayed CBD/MBD E-loop insertions arranged around the two-fold axis of symmetry. Surface representation is shown. MBDS highlighted in pink, CBDs shown in green. MBD: metal-binding domain. CBD: cyclic nucleotide-binding fold domain. (B) A single CBD/MBD insertion shown in ribbon representation. The metal ion coordinated by the MBD is highlighted. (C) Close-up view of the metal-binding loop within the MBD highlighting metal coordination. Residues shown in stick representation. The sequence of the metal binding loop within the *S. griseus* MBD is shown (bottom) and residues involved in metal coordination are highlighted in green. (D) Comparison of the putative CBD ligand binding pocket and the cAMP binding site found in *E. coli* catabolite activator protein (CAP). cAMP (yellow) and key binding site residues are shown in stick representation.

the Flp fusion domains could not be resolved. This was likely caused by the Flp domain being flexibly tethered to the N-terminus of the HK97 phage-like fold, resulting in substantial mobility of internalized Flp domains, thus deviating from crystallographic symmetry. A predicted Alpha Fold structure of the *P. furiosus* protomer highlights that the N-terminal domain consists of a two-helix bundle in isolation (Figure 6B). Recent structural studies of the *P. furiosus* Flp cargo protein alone – with the HK97-type domain removed – show that it forms a D5 symmetrical homo-decameric helical assembly (Figure 6C) which has been suggested to be also present in the natively assembled Flp fusion encapsulin.^[35] To realize decameric Flp assemblies within the context of the T=3 geometry of the *P. furiosus* shell, consisting of twelve pentamers and twenty hexamers, the five Flp domains of a given pentamer and one Flp domain each of the five surrounding hexamers would have to form a decameric Flp complex (Figure 6D). With full engagement at the pentamers, resulting in twelve Flp decamers per shell, sixty Flp domains would remain free. It remains to be seen if this model is correct and if so, what the function of the uncomplexed Flp domains may be.

Another N-terminal fusion domain found in Family 1 encapsulins has previously been computationally identified in the genomes of anammox bacteria of the phylum Planctomycetes.^[36, 86, 91] These unusual encapsulins – called cEnc – carry a diheme cytochrome c fusion domain (Cytc) at their N-terminus, connected to the HK97 phage-like fold via a flexible linker (Figure 6E). Often, a conserved

enzyme component, annotated as a copper reductase-hydroxylamine oxidoreductase fusion protein, can be found right next to the cEnc-encoding gene and may represent a cEnc cargo protein. Anammox bacteria are able to carry out anaerobic ammonium oxidation inside a lipid-based intracellular organelle called the anammoxosome.^[92-94] Proteins targeted to the anammoxosome interior usually carry N-terminal signal peptides necessary for translocation. Such a signal peptide is also found in cEnc protomers, indicating that these unusual encapsulins are localized to the interior of the anammoxosome organelle (Figure 6E). Alpha Fold-based structure prediction of the cEnc protomer suggests that the Cytc domain directly interacts with the P-domain and N-helix of the HK97 phage-like domain (Figure 6F). The diheme Cytc domain is predicted to form a compact globular fold with the two covalently bound heme groups stacked parallel to one another (Figure 6G). Diheme Cytc domains in anammox bacteria are often involved in electron transfer reactions between anammox-related enzymes.^[92, 93] In analogy, Cytc domains within cEnc might be able to transfer electrons across the encapsulin shell. This might be necessary for the activity of any putative cargo enzymes. cEnc has also been suggested to be able to mineralize iron within the anammoxosome through a so far unknown mechanism. Iron-rich electron-dense puncta, too large to be ferritins, are often observed within the anammoxosome in electron microscopy thin sections which could represent cEnc encapsulins.^[93, 95, 96]

REVIEW

Computational searches have identified another type of putative fusion encapsulin classified as a Family 3 system.^[19] These fusion encapsulins carry a C-terminal helical domain (HD) and are found in the genus *Mesorhizobium* (Figure 6H). Genome neighborhood

analysis indicates that they are always embedded within small molecule biosynthetic gene clusters of unknown function. The HD at the C-terminus is predicted to contain four helices with partial

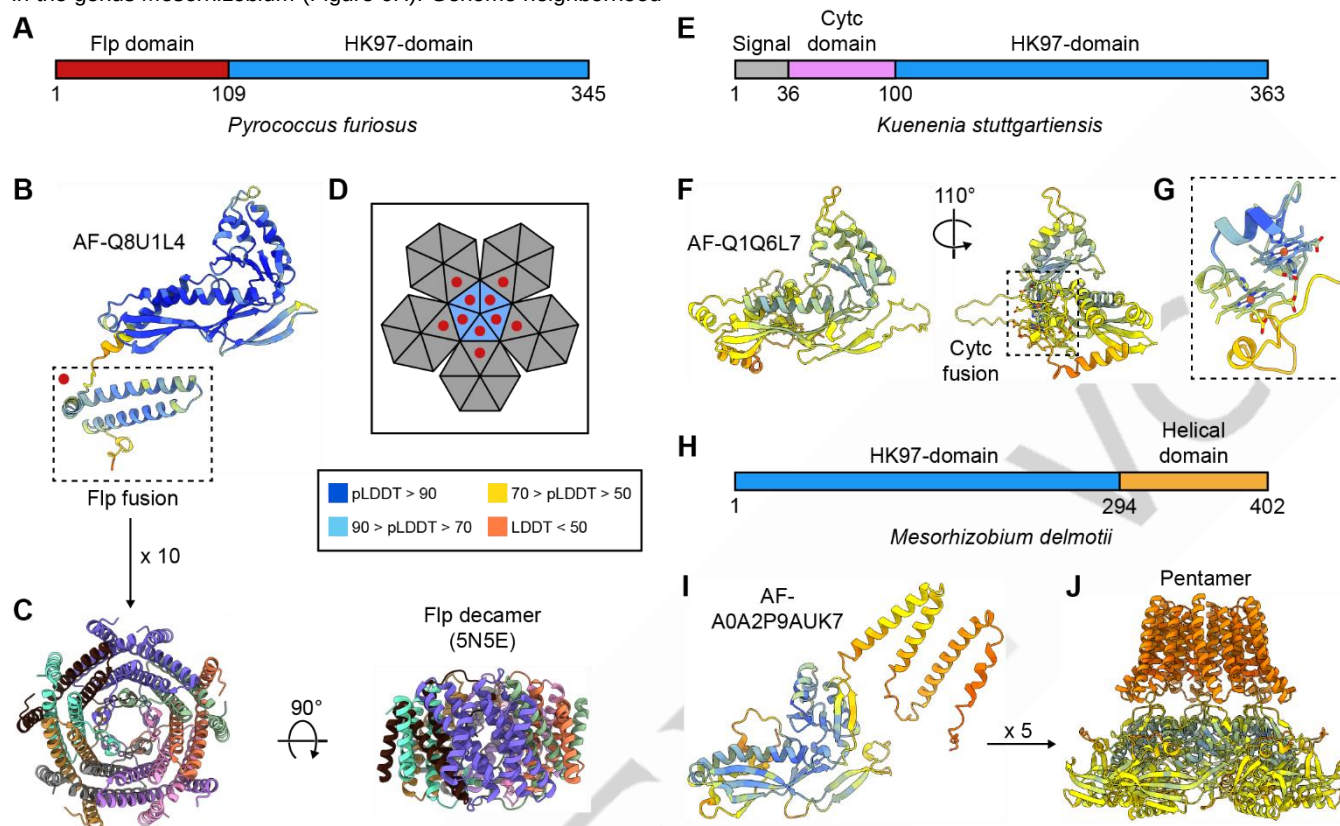


Figure 6. Fusion encapsulin systems. (A) Domain organization of archaeal Family 1 Flp fusion encapsulins. Flp: ferritin-like protein. (B) Alpha Fold prediction of the *P. furiosus* Flp fusion protomer. All predicted structures are colored based on pLDDT scores. (C) Homodecameric helical assembly of the excised *P. furiosus* Flp. (D) Schematic highlighting individual Flps (red dot) involved in the formation of decameric complexes located below the five-fold axis in the context of an assembled T=3 shell. (E) Domain organization of Family 1 cEnc encapsulins found in anammox bacteria. (F) Predicted Alpha Fold structure of a cEnc protomer. (G) Alpha Fold prediction of the diheme cytochrome c fusion domain (CytC) highlighting the two stacked heme groups in stick representation. (H) Domain organization of helical domain (HD)-containing Family 3 encapsulins. (I) Alpha Fold prediction of an HD-containing fusion protomer. (J) Alpha Fold-predicted pentamer highlighting the formation of a helical bundle above the five-fold pore.

membrane helix character, containing a substantial number of hydrophobic residues (Figure 6I). Considering that the C-terminus within the HK97 phage-like fold protrudes to the shell exterior and is located close to the symmetry axes and that within the context of an icosahedral encapsulin shell, pentamers and hexamers could be formed, five or six helical domains may interact with one another. Indeed, an Alpha Fold prediction of five copies of the *Mesorhizobium delmotii* protomer yields an assembled pentamer with HDs forming an external channel-like protrusion located on top of the five-fold pore (Figure 6J). Similar structures are predicted for hexameric assemblies. Within a putative pentamer, the exterior helical channel consists of twenty helices with the outside of the helical bundle predicted to exhibit substantial surface hydrophobicity. While no experimental data is available for these systems, these structure predictions may suggest that HD-containing Family 3 encapsulins are able to interact with lipid membranes and that their HDs can act as channels for transmitting small molecules into and out of the encapsulin shell.

5. Summary and Outlook

While substantial progress has been made over the past five years with respect to the structural understanding of encapsulin assemblies, many challenges remain. One of them is to elucidate the dynamics of encapsulin pores and the triggers that favor one pore state over another. This includes both Family 1 five-fold and Family 2 two-fold pores, and potentially other types of pores. Advanced computational tools for the analysis of discrete and continuous heterogeneity in cryo-electron microscopy data are now available and will likely be indispensable for addressing these questions. The dynamics of CBD/MBD insertions and their role in determining the two-fold pore state in Family 2B shells is currently not well understood and could likewise be addressed through detailed analyses of CBD/MBD conformational flexibility. In addition, the identification of CBD ligands and how ligand binding influences Family 2B shell conformation and biological function should be prioritized in the future. The recent discovery of two-component Family 2B encapsulin shells further adds to the complexity of these assemblies and detailed future studies are needed to better understand their assembly constraints and structural states. While information about the assembly mechanisms of some HK97-type viral capsids is currently

REVIEW

available, how individual encapsulin protomers associate and self-assemble into closed icosahedral shells is still unknown. Progress could be made through *in vitro* studies of engineered protomers or *in vitro* disassembly and reassembly assays using pH or ionic strength shifts. Ideally though, shell assembly could be followed *in situ* using cryo-electron tomography to capture native assembly intermediates.

Currently, no or limited structural information is available for Family 3 and Family 4 encapsulins. Future efforts should be focused on investigating these systems in general and determining their assembly states and structures in particular. As most of these uncharacterized encapsulins are found in high GC Gram-positive organisms (Family 3) or hyperthermophilic and anaerobic archaea (Family 4), innovative heterologous expression or native isolation strategies might need to be employed to produce these systems in native-like states for downstream characterization. Similarly, as no structural information for cEnc and HD fusion encapsulins is available at the moment, their structural characterization should also be prioritized.

Basic studies investigating encapsulin structure, self-assembly, and cargo loading have so far provided a tremendously useful molecular toolbox for encapsulin engineering aimed at utilizing encapsulin-based systems for applications in biomedicine, biocatalysis, and bionanotechnology. It is likely that continued discovery- and curiosity-driven work on encapsulins will expand this toolbox and lead to novel innovative and impactful synthetic biology use cases of engineered encapsulins with real-world impact.

Acknowledgements

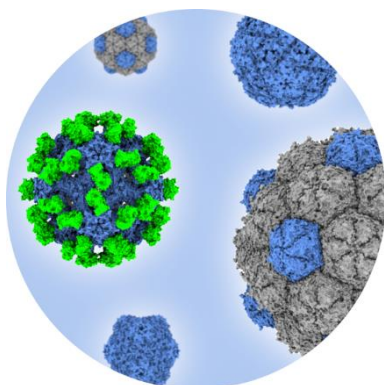
I acknowledge funding from the NIH (R35GM133325), NSF (2342136), and the UM Research Scouts program (OORRS033123). Molecular graphics analyses were performed using UCSF ChimeraX developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from the National Institutes of Health R01GM129325 and the Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases.

Keywords: Encapsulin • Nanocompartment • Protein shell • Protein capsid • HK97

- [1] E. Cornejo, N. Abreu, A. Komeili, *Curr Opin Cell Biol* **2014**, 26, 132-138.
- [2] Y. Diekmann, J. B. Pereira-Leal, *Biochem J* **2013**, 449, 319-331.
- [3] T. Gabaldón, A. A. Pitts, *Biochimie* **2015**, 119, 262-268.
- [4] C. Aussignargues, M. E. Pandelia, M. Sutter, J. S. Plegaria, J. Zarzycki, A. Turmo, J. Huang, D. C. Ducat, E. L. Hegg, B. R. Gibney, C. A. Kerfeld, *J Am Chem Soc* **2016**, 138, 5262-5270.
- [5] C. A. Kerfeld, S. Heinhorst, G. C. Cannon, *Annu Rev Microbiol* **2010**, 64, 391-408.
- [6] C. Greening, T. Lithgow, *Nat Rev Microbiol* **2020**.
- [7] L. S. R. Adamson, N. Tasneem, M. P. Andreas, W. Close, E. N. Jenner, T. N. Szyszka, R. Young, L. C. Cheah, A. Norman, H. I. MacDermott-Opeskin, M. L. O'Mara, F. Sainsbury, T. W. Giessen, Y. H. Lau, *Sci Adv* **2022**, 8, eabl7346.
- [8] T. W. Giessen, *Annu Rev Biochem* **2022**, 91, 353-380.
- [9] R. J. Nichols, C. Cassidy-Amstutz, T. Chaijarasphong, D. F. Savage, *Crit Rev Biochem Mol Biol* **2017**, 52, 583-594.
- [10] T. W. Giessen, P. A. Silver, *J Mol Biol* **2016**, 428, 916-927.
- [11] S. Govindarajan, K. Nevo-Dinur, O. Amster-Choder, *FEMS Microbiol Rev* **2012**, 36, 1005-1022.
- [12] M. R. Melnicki, M. Sutter, C. A. Kerfeld, *Curr Opin Microbiol* **2021**, 63, 1-9.
- [13] J. M. Ochoa, T. O. Yeates, *Curr Opin Microbiol* **2021**, 62, 51-60.
- [14] A. Turmo, C. R. Gonzalez-Esquer, C. A. Kerfeld, *FEMS Microbiol Lett* **2017**, 364.
- [15] L. N. Liu, *Trends Microbiol* **2021**.
- [16] T. A. Bobik, B. P. Lehman, T. O. Yeates, *Mol Microbiol* **2015**, 98, 193-207.
- [17] B. Ferlez, M. Sutter, C. A. Kerfeld, *mBio* **2019**, 10.
- [18] M. B. Prentice, *Curr Opin Microbiol* **2021**, 63, 19-28.
- [19] M. P. Andreas, T. W. Giessen, *Nat Commun* **2021**, 12, 4748.
- [20] M. Sutter, D. Boehringer, S. Gutmann, S. Günther, D. Prangishvili, M. J. Loessner, K. O. Stetter, E. Weber-Ban, N. Ban, M. Sutter, D. Boehringer, S. Gutmann, S. Günther, D. Prangishvili, M. J. Loessner, K. O. Stetter, E. Weber-Ban, N. Ban, *Nature Structural & Molecular Biology* **2008** 15:9 **2008-08-31**, 15.
- [21] F. Akita, K. T. Chong, H. Tanaka, E. Yamashita, N. Miyazaki, Y. Nakaishi, M. Suzuki, K. Namba, Y. Ono, T. Tsukihara, A. Nakagawa, *J Mol Biol* **2007**, 368, 1469-1483.
- [22] T. W. Giessen, B. J. Orlando, A. A. Verdegaaal, M. G. Chambers, J. Gardener, D. C. Bell, G. Birrane, M. Liao, P. A. Silver, *Elife* **2019**, 8.
- [23] J. A. Jones, R. Benisch, T. W. Giessen, *J Mater Chem B* **2023**, 11, 4377-4388.
- [24] W. J. Altenburg, N. Rollins, P. A. Silver, T. W. Giessen, *Sci Rep* **2021**, 11, 4951.
- [25] J. A. Jones, M. P. Andreas, T. W. Giessen, *Nat Commun* **2024**, 15, 2558.
- [26] S. Kwon, M. P. Andreas, T. W. Giessen, *J Struct Biol* **2023**, 215, 108022.
- [27] R. Benisch, M. P. Andreas, T. W. Giessen, *Sci Adv* **2024**, 10, eadk9345.
- [28] M. P. Andreas, T. W. Giessen, *bioRxiv* **2024**, 2024.2004.2023.590730.
- [29] R. J. Nichols, B. LaFrance, N. R. Phillips, D. R. Radford, L. M. Oltrogge, L. E. Valentin-Alvarado, A. J. Bischoff, E. Nogales, D. F. Savage, *Elife* **2021**, 10.
- [30] K. A. Lien, K. Dinshaw, R. J. Nichols, C. Cassidy-Amstutz, M. Knight, R. Singh, L. D. Eltis, D. F. Savage, S. A. Stanley, *Elife* **2021**, 10.
- [31] R. Rahmanpour, T. D. Bugg, *FEBS J* **2013**, 280, 2097-2104.
- [32] Y. Tang, A. Mu, Y. Zhang, S. Zhou, W. Wang, Y. Lai, X. Zhou, F. Liu, X. Yang, H. Gong, Q. Wang, Z. Rao, *Proc Natl Acad Sci U S A* **2021**, 118.
- [33] C. A. McHugh, J. Fontana, D. Nemecek, N. Cheng, A. A. Aksyuk, J. B. Heymann, D. C. Winkler, A. S. Lam, J. S. Wall, A. C. Steven, E. Hoiczky, *EMBO J* **2014**, 33, 1896-1911.
- [34] D. He, S. Hughes, S. Vanden-Hehir, A. Georgiev, K. Altenbach, E. Tarrant, C. L. Mackay, K. J. Waldron, D. J. Clarke, J. Marles-Wright, *Elife* **2016**, 5.
- [35] D. He, C. Piergentili, J. Ross, E. Tarrant, L. R. Tuck, C. L. Mackay, Z. McIver, K. J. Waldron, D. J. Clarke, J. Marles-Wright, *Biochem J* **2019**, 476, 975-989.
- [36] M. P. Andreas, T. W. Giessen, *MethodsX* **2022**, 9, 101787.
- [37] T. Wiryaman, N. Toor, *J Struct Biol X* **2022**, 6, 100062.
- [38] N. S. Chmelyuk, V. V. Oda, A. N. Gabashvili, M. A. Abakumov, *Biochemistry (Mosc)* **2023**, 88, 35-49.
- [39] A. R. Quinton, H. B. McDowell, E. Hoiczky, *Adv Appl Microbiol* **2023**, 125, 1-48.
- [40] H. B. McDowell, E. Hoiczky, *J Bacteriol* **2022**, 204, e0034621.

REVIEW

- [41] J. M. Rodriguez, C. Allende-Ballester, J. Cornelissen, J. R. Caston, *Nanomaterials (Basel)* **2021**, *11*.
- [42] J. A. Jones, T. W. Giessen, *Biotechnol Bioeng* **2021**, *118*, 491-505.
- [43] A. N. Gabashvili, N. S. Chmelyuk, M. V. Efremova, J. A. Malinovskaya, A. S. Semkina, M. A. Abakumov, *Biomolecules* **2020**, *10*.
- [44] A. M. Demchuk, T. R. Patel, *Biotechnol Adv* **2020**, *41*, 107547.
- [45] T. W. Giessen, P. A. Silver, *Curr Opin Biotechnol* **2017**, *46*, 42-50.
- [46] T. W. Giessen, *Curr Opin Chem Biol* **2016**, *34*, 1-10.
- [47] J. B. Shabb, J. D. Corbin, *J Biol Chem* **1992**, *267*, 5723-5726.
- [48] E. Eren, B. Wang, D. C. Winkler, N. R. Watts, A. C. Steven, P. T. Wingfield, *Structure* **2022**, *30*, 551-563 e554.
- [49] C. A. Dutcher, M. P. Andreas, T. W. Giessen, *bioRxiv* **2024**.
- [50] J. Abramson, J. Adler, J. Dunger, R. Evans, T. Green, A. Pritzel, O. Ronneberger, L. Willmore, A. J. Ballard, J. Bambrick, S. W. Bodenstern, D. A. Evans, C. C. Hung, M. O'Neill, D. Reiman, K. Tunyasuvunakool, Z. Wu, A. Zemgulyte, E. Arvaniti, C. Beattie, O. Bertolli, A. Bridgland, A. Cherepanov, M. Congreve, A. I. Cowen-Rivers, A. Cowie, M. Figurnov, F. B. Fuchs, H. Gladman, R. Jain, Y. A. Khan, C. M. R. Low, K. Perlin, A. Potapenko, P. Savy, S. Singh, A. Stecula, A. Thillaisundaram, C. Tong, S. Yakneen, E. D. Zhong, M. Zielinski, A. Zidek, V. Bapst, P. Kohli, M. Jaderberg, D. Hassabis, J. M. Jumper, *Nature* **2024**, *630*, 493-500.
- [51] J. Mistry, S. Chuguransky, L. Williams, M. Qureshi, G. A. Salazar, E. L. L. Sonhammer, S. C. E. Tosatto, L. Paladin, S. Raj, L. J. Richardson, R. D. Finn, A. Bateman, *Nucleic Acids Res* **2021**, *49*, D412-D419.
- [52] Y. Liu, T. A. Demina, S. Roux, P. Aiewsakun, D. Kazlauskas, P. Simmonds, D. Prangishvili, H. M. Oksanen, M. Krupovic, *PLoS Biol* **2021**, *19*, e3001442.
- [53] W. R. Wikoff, L. Liljas, R. L. Duda, H. Tsuruta, R. W. Hendrix, J. E. Johnson, *Science* **2000**, *289*, 2129-2133.
- [54] M. K. Pietilä, P. Laurinmäki, D. A. Russell, C.-C. Ko, D. Jacobs-Sera, R. W. Hendrix, D. H. Bamford, S. J. Butcher, M. K. Pietilä, P. Laurinmäki, D. A. Russell, C.-C. Ko, D. Jacobs-Sera, R. W. Hendrix, D. H. Bamford, S. J. Butcher, *Proceedings of the National Academy of Sciences* **2013-6-25**, *110*.
- [55] M. L. Baker, W. Jiang, F. J. Rixon, W. Chiu, *J Virol* **2005**, *79*, 14967-14970.
- [56] M. M. Suhanovsky, C. M. Teschke, *Virology* **2015**, *479-480*, 487-497.
- [57] Z. H. Zhou, W. H. Hui, S. Shah, J. Jih, C. M. O'Connor, M. B. Sherman, D. H. Kedes, S. Schein, *Structure* **2014**, *22*, 1385-1398.
- [58] D. L. CASPAR, A. KLUG, *Cold Spring Harb Symp Quant Biol* **1962**, *27*, 1-24.
- [59] J. E. Johnson, J. A. Speir, *J Mol Biol* **1997**, *269*, 665-675.
- [60] T. Wiryaman, N. Toor, *IUCrJ* **2021**, *8*, 342-350.
- [61] B. J. LaFrance, C. Cassidy-Amstutz, R. J. Nichols, L. M. Oltrogge, E. Nogales, D. F. Savage, *Sci Rep* **2021**, *11*, 22810.
- [62] R. Zinelli, S. Soni, J. Cornelissen, S. Michel-Souzy, C. A. Nijhuis, *Biomolecules* **2023**, *13*.
- [63] A. Fokine, P. G. Leiman, M. M. Shneider, B. Ahvazi, K. M. Boeshans, A. C. Steven, L. W. Black, V. V. Mesyanzhinov, M. G. Rossmann, *Proc Natl Acad Sci U S A* **2005**, *102*, 7163-7168.
- [64] M. L. Yap, M. G. Rossmann, *Future Microbiol* **2014**, *9*, 1319-1327.
- [65] J. Ross, Z. McIver, T. Lambert, C. Piergentili, J. E. Bird, K. J. Gallagher, F. L. Cruickshank, P. James, E. Zarazua-Arzu, L. E. Horsfall, K. J. Waldron, M. D. Wilson, C. L. Mackay, A. Basle, D. J. Clarke, J. Marles-Wright, *Sci Adv* **2022**, *8*, eabj4461.
- [66] N. Loncar, H. J. Rozeboom, L. E. Franken, M. C. A. Stuart, M. W. Fraaije, *Biochem Biophys Res Commun* **2020**, *529*, 548-553.
- [67] J. A. Jones, M. P. Andreas, T. W. Giessen, *Biomacromolecules* **2023**, *24*, 1388-1399.
- [68] S. Kwon, M. P. Andreas, T. W. Giessen, *bioRxiv* **2024**.
- [69] E. M. Williams, S. M. Jung, J. L. Coffman, S. Lutz, *ACS Synth Biol* **2018**, *7*, 2514-2517.
- [70] M. C. Jenkins, S. Lutz, *ACS Synth Biol* **2021**, *10*, 857-869.
- [71] T. N. Szyszka, M. P. Andreas, F. Lie, L. M. Miller, L. S. R. Adamson, F. Fatehi, R. Twarock, B. E. Draper, M. F. Jarrold, T. W. Giessen, Y. H. Lau, *Proc Natl Acad Sci U S A* **2024**, *121*, e2321260121.
- [72] R. L. Duda, C. M. Teschke, *Curr Opin Virol* **2019**, *36*, 9-16.
- [73] A. Fokine, P. R. Chipman, P. G. Leiman, V. V. Mesyanzhinov, V. B. Rao, M. G. Rossmann, *Proc Natl Acad Sci U S A* **2004**, *101*, 6003-6008.
- [74] M. C. Morais, K. H. Choi, J. S. Koti, P. R. Chipman, D. L. Anderson, M. G. Rossmann, *Mol Cell* **2005**, *18*, 149-159.
- [75] K. N. Parent, R. Khayat, L. H. Tu, M. M. Suhanovsky, J. R. Cortines, C. M. Teschke, J. E. Johnson, T. S. Baker, *Structure* **2010**, *18*, 390-401.
- [76] D. H. Chen, M. L. Baker, C. F. Hryc, F. DiMaio, J. Jakana, W. Wu, M. Dougherty, C. Haase-Pettingell, M. F. Schmid, W. Jiang, D. Baker, J. A. King, W. Chiu, *Proc Natl Acad Sci U S A* **2011**, *108*, 1355-1360.
- [77] K. N. Parent, E. B. Gilcrease, S. R. Casjens, T. S. Baker, *Virology* **2012**, *427*, 177-188.
- [78] K. N. Parent, J. Tang, G. Cardone, E. B. Gilcrease, M. E. Janssen, N. H. Olson, S. R. Casjens, T. S. Baker, *Virology* **2014**, *464-465*, 55-66.
- [79] F. L. Homa, J. B. Huffman, K. Toropova, H. R. Lopez, A. M. Makhov, J. F. Conway, *J Mol Biol* **2013**, *425*, 3415-3428.
- [80] W. H. Hui, Q. Tang, H. Liu, I. Atanasov, F. Liu, H. Zhu, Z. H. Zhou, *Protein Cell* **2013**, *4*, 833-845.
- [81] Z. H. Zhou, D. H. Chen, J. Jakana, F. J. Rixon, W. Chiu, *J Virol* **1999**, *73*, 3210-3218.
- [82] Z. H. Zhou, M. Dougherty, J. Jakana, J. He, F. J. Rixon, W. Chiu, *Science* **2000**, *288*, 877-880.
- [83] X. Dai, Z. H. Zhou, *Science* **2018**, *360*.
- [84] R. L. Duda, K. Martincic, R. W. Hendrix, *J Mol Biol* **1995**, *247*, 636-647.
- [85] G. Effantin, P. Boulanger, E. Neumann, L. Letellier, J. F. Conway, *J Mol Biol* **2006**, *361*, 993-1002.
- [86] T. W. Giessen, P. A. Silver, *Nat Microbiol* **2017**, *2*, 17029.
- [87] H. M. Berman, L. F. Ten Eyck, D. S. Goodsell, N. M. Haste, A. Kornev, S. S. Taylor, *Proc Natl Acad Sci U S A* **2005**, *102*, 45-50.
- [88] N. Kannan, J. Wu, G. S. Anand, S. Yooseph, A. F. Neuwald, J. C. Venter, S. S. Taylor, *Genome Biol* **2007**, *8*, R264.
- [89] J. M. Passner, S. C. Schultz, T. A. Steitz, *J Mol Biol* **2000**, *304*, 847-859.
- [90] N. Popovich, S. R. Tzeng, M. Tonelli, R. H. Ebright, C. G. Kalodimos, *Proc Natl Acad Sci U S A* **2009**, *106*, 6927-6932.
- [91] J. C. Tracey, M. Coronado, T. W. Giessen, M. C. Y. Lau, P. A. Silver, B. B. Ward, *Sci Rep* **2019**, *9*, 20122.
- [92] B. Kartal, J. T. Keltjens, *Trends Biochem Sci* **2016**, *41*, 998-1011.
- [93] L. van Niftrik, *Antonie Van Leeuwenhoek* **2013**, *104*, 489-497.
- [94] J. G. Kuenen, *Nat Rev Microbiol* **2008**, *6*, 320-326.
- [95] L. van Niftrik, W. J. Geerts, E. G. van Donselaar, B. M. Humbel, A. Yakushevskaya, A. J. Verkleij, M. S. Jetten, M. Strous, *J Struct Biol* **2008**, *161*, 401-410.
- [96] M. Strous, E. Pelletier, S. Mangenot, T. Rattei, A. Lehner, M. W. Taylor, M. Horn, H. Daims, D. Bartol-Mavel, P. Wincker, V. Barbe, N. Fonknechten, D. Vallenet, B. Segurens, C. Schenowitz-Truong, C. Medigue, A. Collingro, B. Snel, B. E. Dutilh, H. J. Op den Camp, C. van der Drift, I. Cirpus, K. T. van de Pas-Schoonen, H. R. Harhangi, L. van Niftrik, M. Schmid, J. Keltjens, J. van de Vossenberg, B. Kartal, H. Meier, D. Frishman, M. A. Huynen, H. W. Mewes, J. Weissenbach, M. S. Jetten, M. Wagner, D. Le Paslier, *Nature* **2006**, *440*, 790-794.

Entry for the Table of Contents

Encapsulin nanocompartments are one of the most diverse and widespread classes of prokaryotic protein compartments. This review focuses on the structural diversity of encapsulins with a focus on shell structure, pore properties, and domain insertions and fusions found within the encapsulin shell protein.

Institute and/or researcher Twitter usernames: @GiessenLab