

Characterization of a Primordial Major Capsid-Scaffolding Protein Complex in Icosahedral Virus Shell Assembly

Christal R. Davis¹, Donald Backos², Marc C. Morais³, Mair E. A. Churchill^{1,4} and Carlos E. Catalano^{1,2*}

1 - Program in Structural Biology and Biochemistry, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

2 - Department of Pharmaceutical Chemistry, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

3 - Department of Biochemistry and Molecular Biology, University of Texas Medical Branch, Galveston, TX, USA

4 - Department of Pharmacology, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

Correspondence to Carlos E. Catalano: Skaggs School of Pharmacy and Pharmaceutical Sciences, Department of Pharmaceutical Sciences, University of Colorado, Anschutz Medical Campus, 12850 Montview Ave., Aurora, CO 80045, USA. carlos.catalano@cuanschutz.edu (C.E. Catalano)

<https://doi.org/10.1016/j.jmb.2022.167719>

Edited by Owen Pornillos

Abstract

Capsid assembly pathways are strongly conserved in the complex dsDNA viruses, where major capsid proteins (MCP) self-assemble into icosahedral procapsid shells, chaperoned by a scaffolding protein. Without a scaffold, the capsid proteins aggregate and form aberrant structures. This, coupled with the rapid co-polymerization of MCP and scaffolding proteins, has thwarted characterization of the earliest steps in shell assembly. Here we interrogate the structure and biophysical properties of a soluble, assembly-deficient phage lambda major capsid protein, **MCP(W308A)**. The mutant protein is folded, soluble to high concentrations and binds to the scaffolding protein in an apparent $SP_2:MCP(W308A)_1$ stoichiometry but does not assemble beyond this initiating complex. The MCP(W308A) crystal structure was solved to 2.7 Å revealing the canonical HK97 fold in a “pre-assembly” conformation featuring the conserved N-arm and E-loops folded into the body of the protein. Structural, biophysical and computational analyses suggest that MCP(W308A) is thermodynamically trapped in this pre-assembly conformation precluding self-association interactions required for shell assembly. A model is described wherein dynamic interactions between MCP proteins play an essential role in high fidelity viral shell assembly. Scaffold-chaperoned MCP polymerization is a strongly conserved process in all the large dsDNA viruses and our results provide insight into this primordial complex in solution and have broad biological significance in our understanding of virus assembly mechanisms.

© 2022 Elsevier Ltd. All rights reserved.

Introduction

All viruses protect their genomes by encapsulating them in a protein coat.¹ The complex double-stranded DNA (dsDNA) viruses, which include the caudoviruses (tailed bacteriophages) and the eukaryotic herpesvirus groups, possess

conserved developmental pathways^{2–4}; viral DNA is typically replicated via a rolling circle mechanism to afford linear concatemers of viral genomes linked in a head to tail fashion. Separately, capsid proteins are expressed, and icosahedral *procapsid* shells self-assemble by co-polymerization of one or a few major capsid proteins and a scaffolding protein,

which functions as a molecular chaperone to dictate the final size and shape of the shell.^{5–7} In the absence of a scaffolding protein, the major capsid proteins polymerize into aberrant structures and insoluble aggregates both *in vivo*^{8–9} and *in vitro*.^{10–12} Packaging of viral DNA into the pre-assembled procapsids is catalyzed by a terminase enzyme and represents the intersection of the DNA replication and procapsid assembly pathways. Genome packaging triggers a remarkable, large-scale reorganization of the procapsid shell to afford a mature *capsid* structure. Upon insertion of the full-length genome, terminase is released from the nucleocapsid and addition of a tail (caudoviruses) or a lipid envelope (herpesviruses) affords an infectious virus.

Procapsid assembly, as outlined above, is strongly conserved in the complex dsDNA viruses which affords icosahedral shells composed of the major capsid protein(s) (MCP) and the scaffolding protein (SP) that resides at the shell interior.^{3–4,6,10,13–18} The scaffolding protein exits the shell prior to, or concomitant with DNA packaging *in vivo*. In some cases, the SP exits intact and is recycled (e.g., phages P22, ϕ 29) while in others it is degraded by a viral protease and the fragments exit the shell (e.g., phage λ , the herpesviruses). In both cases, SP release from the assembled shell make accurate determination of the SP:MCP stoichiometry difficult even in the earliest procapsid structures. This complicates mechanistic interrogation of the molecular interactions driving co-polymerization of the two proteins.

The interaction interfaces between SP and MCP in fully assembled shells has been described in a number of cryo-EM structures. These studies reveal that the scaffolding proteins assemble as dimers at the shell interior.^{19–22} Further, C-terminal SP residues typically interact with N-terminal residues of the capsid proteins in a 1:1 or 1:2 stoichiometry (SP:MCP). Thus, while structural studies have provided insight into protein–protein interactions in the assembled shells, the rapid co-polymerization of scaffolding and capsid proteins in solution, the transient nature of their

interaction and the propensity of the isolated capsid proteins to aggregate has thwarted direct interrogation of the SP-MCP interactions responsible for nucleating high fidelity shell assembly.

Phage λ is representative of the complex dsDNA viruses and provides as an excellent model system in which to study virus assembly.^{5,17,23} As with all these viruses, the λ procapsid self-assembles by co-polymerization of a major capsid protein (gpE, MCP) and a scaffolding protein (gpNu3, SP) to afford an icosahedral shell (Figure 1). Three MCP hexamers form each of the 20 faces of the icosahedron while pentamers reside at eleven of the vertices; the twelfth vertex contains the genome packaging portal protein, which provides a conduit for DNA insertion during packaging and genome egress during cell infection. The hexamers and pentamers, which are collectively known as *capsomers*, converge at the threefold axes of the assembled shell, which contains 70–300 copies of SP.^{24–26} Structural models²⁷ and recent cryo-EM structures of λ procapsid and capsid shells²⁸ reveal that the MCP retains the conserved HK97 fold found in dsDNA phages and the herpesviruses, including the A-domain and the P-domain featuring the prominent “spine helix”.

We previously reported that while purified λ MCP aggregates in solution, soluble λ procapsids can be assembled in the presence of SP *in vitro*.²⁹ Subsequently, we described a mutant λ MCP protein, **MCP(W308A)**, in which Trp308 was changed to Ala. Unlike the wild-type protein, this variant remains a soluble monomer up to 80 μ M.³⁰ Moreover, while MCP(W308A) retains SP binding interactions, it does not assemble into higher-order complexes. This mutant provides an unprecedented opportunity to interrogate elemental protein–protein interactions that lead to procapsid assembly, and here we report biophysical, computational and structural studies that characterize MCP(W308A) and its interaction with SP in solution. The study provides insight into the fundamental interactions that initiate high-fidelity procapsid shell assembly, a process that is strongly conserved in all of the complex dsDNA viruses.

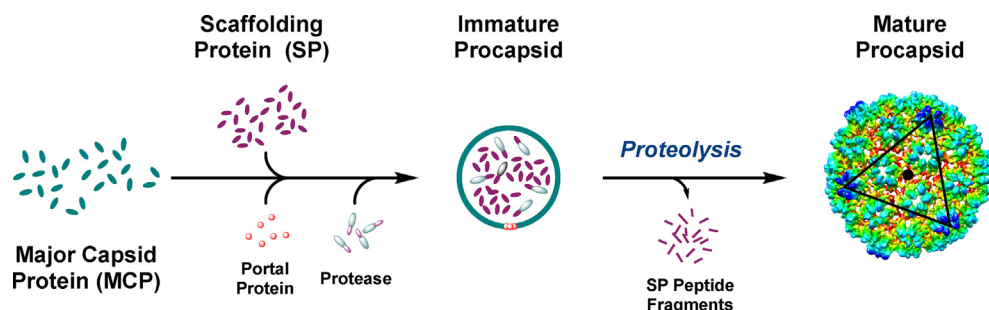


Figure 1. Lambda Capsid Assembly *in vivo*. Triangle denotes an icosahedral face composed of three MCP hexamers with the threefold shell axis indicated with a black dot. Details are provided in the text.

Results

MCP(W308A) undergoes a monomer-dimer self-association interaction

As discussed above, major capsid proteins naturally self-associate, forming aberrant, non-capsid structures and aggregates in the absence of the scaffolding protein.^{10–12,29,31} In contrast, we previously reported that a mutant λ capsid protein, MCP(W308A), remains a soluble monomer up to 80 μM as determined by size exclusion chromatography (SEC).³⁰ Here we rigorously interrogate MCP(W308A) self-association using sedimentation velocity analytical ultracentrifugation (SV-AUC). At a concentration of 20 μM , MCP(W308A) sediments as a single species and analysis of the data using SedFit returns a molecular weight of 36.5 kDa (Figure 2A). This is consistent with the monomer molecular weight determined from the gene sequence

(38.0 kDa) and indicates that MCP(W308A) is a monomer at this concentration. The figure further shows that the $s_{(20,w)}$ increases with increasing protein concentration to 200 μM , indicating that the protein self-associates in this concentration range. Accurate determination of the K_D and the stoichiometry for the interaction requires that higher protein concentrations be examined, which is complicated by aggregation of MCP(W308A) at concentrations greater than 200 μM (not shown). We note, however, that association of a monomer to a dimer species results in an increase in the sedimentation coefficient by 1.6-fold.^{32–33} Close inspection of the data reveal that the $s_{(20,w)}$ for MCP(W308A) increases by ~ 1.1 -fold between 50 μM to 200 μM (Figure 2A, inset), which is consistent with a monomer-dimer self-association interaction within this concentration range. Analysis of the data according to a monomer-dimer equilibrium model affords an

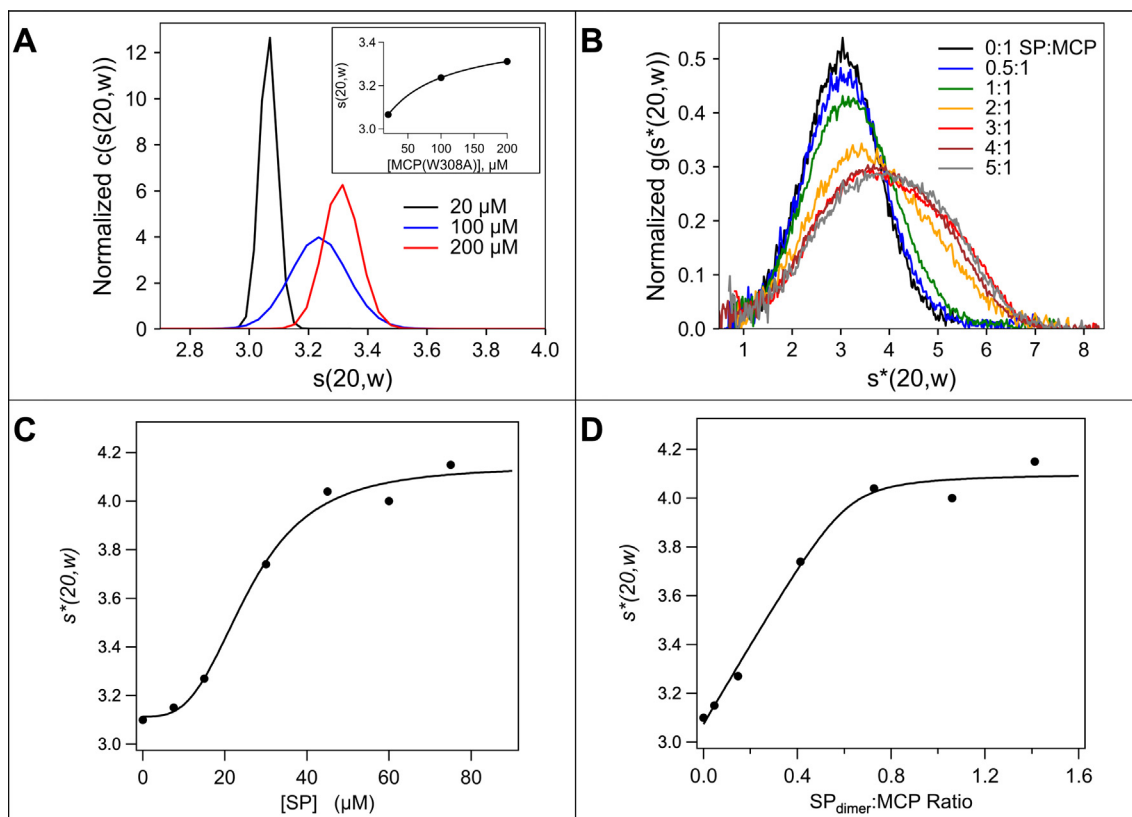


Figure 2. Biophysical Characterization of MCP(W308A) and Its Interaction with SP. *Panel A.* Self-assembly of MCP(W308A) monitored by SV-AUC. Purified MCP(W308A) at the indicated concentration was characterized by SV-AUC and analyzed using SedFit as described in Materials and Methods. *Inset:* Plot of s vs $[\text{MCP}]$. The solid line represents the best fit of the data to a simple Langmuir binding model. *Panel B.* AUC-monitored SP binding to MCP(W308A). SP was incrementally added to 15 μM MCP(W308A) and the complex characterized by SV-AUC and analyzed using DCDT⁺ as described in Materials and Methods. *Panel C.* Quantitation of the data in *Panel B*. The solid line represents the best fit of the data to a phenomenological Hill-Langmuir binding model. *Panel D.* The data in *Panel C* were converted to $[\text{SP}_{\text{dimer}}]$ based on the dimerization $K_{D,\text{app}} = 50 \mu\text{M}$.²⁹ The solid line represents the best fit of the data according to a stoichiometric binding model.

estimation of $K_{D,app} \sim 100 \mu\text{M}$. The concentration of MCP during a productive infection *in vivo* is 70–150 μM ³⁴ and assuming that the wild-type MCP (MCP-WT) behaves similarly, the observed monomer–dimer association is biologically relevant.

MCP(W308A) retains SP binding interactions

We previously used SEC to demonstrate that MCP(W308A) retains SP binding activity, but that this interaction does not promote shell assembly.³⁰ Here we employ SV-AUC to further characterize this interaction. In this experiment, increasing concentrations of SP were added to MCP(W308A) (15 μM) and the binding data were analyzed as described in Materials and Methods. We note that SP does not contain any aromatic residues and the sedimentation profile (monitored at A280) reflects increasing mass of the mutant capsid protein. The SV-AUC data were analyzed using a model independent $g(s)$ approach and the $g(s^*)$ distribution for MCP(W308A) in the absence of SP affords $s^*_{(20,w)} = 3.1 \text{ S}$ (Figure 2B), consistent with the $c(s)$ data presented in Figure 2A. Incremental addition of SP to MCP(W308A) results in concentration-dependent shift of the $g(s^*)$ distribution to larger values (Figure 2B), clearly indicating that the two proteins interact. A plot of $s^*_{(20,w)}$ vs [SP] is shown in Figure 2C and analysis of the data according to a phenomenological Hill–Langmuir binding model affords a $K_{D,app} = 26 \pm 2 \mu\text{M}$. The $s^*_{(20,w)}$ at saturation ($4.1 \pm 0.1 \text{ S}$) indicates that the complex is relatively small, consistent with the failure of SP to promote MCP polymerization. We calculated theoretical $s(20,w)$ values for a series of $\text{SP}_x \bullet \text{MCP(W308A)}_y$ hetero-oligomers within a range of reasonable frictional coefficients (Table S1). Comparison of these values with the experimental data suggest that the 4.1 S species is most likely composed of one or two SP monomers bound to one MCP protein.

Within this context, we previously demonstrated that the λ scaffolding protein undergoes a monomer–dimer self-association reaction in the concentration range used in this study ($K_{D,app} \sim 50 \mu\text{M}$)²⁹. Scaffolding protein dimers represent the biologically active assembly state in phages P22, SPP1, $\phi 29$ ^{21,35–36} and in the herpesviruses,^{37–39} and we presume that this conserved feature is also observed in phage λ .²⁹ Interestingly, a plot of $s^*_{(20,w)}$ vs [SP]_{dimer} (calculated based on $K_{D,app} = 50 \mu\text{M}$) reveals an apparent stoichiometric binding interaction (Figure 2D). Analysis of the data according to a stoichiometric binding model as described in Materials and Methods yields a stoichiometric breakpoint of $0.6 \pm 0.1 \text{ SP}_{\text{dimers}}:\text{MCP}$, or one SP dimer bound to one or possibly-two MCPs, on average. We note that while an accurate estimate of the binding constant is error prone in the analysis of stoichiometric binding data, the K_D for the interaction is an order of magnitude lower than the concentration of the macro-

molecule,⁴⁰ 15 μM in this case. Thus, we presume that $K_{D,app} < 1.5 \mu\text{M}$.

We next utilized CD spectroscopy to monitor binding of the scaffolding protein to MCP(W308A). Incremental addition of SP to MCP(W308A) does not alter the *far*-UV CD spectrum of either protein (data not shown), indicating that secondary structural elements do not change in the nucleation complex. We next used *near*-UV CD spectroscopy to monitor the binding interaction. As anticipated for a protein devoid of aromatic residues, the spectrum for the SP alone is flat (Figure S1). In contrast, the spectrum for MCP(W308A) is rich in fine structure, including prominent bands at 260–270 nm, 270–280 nm and 280–300 nm, most likely originating from Phe, Tyr and Trp residues, respectively. Titration of scaffolding protein into MCP(W308A) little affects the dominant 292 nm band, which arises from the two Trp residues in the protein,²⁷ suggesting that SP binds distant from these residues. In contrast, significant changes are observed in the bands centered at 276 nm and 284 nm, and to a lesser extent to the 266 nm bands. This again demonstrates an interaction between the two proteins, and further indicates that SP binds to MCP(W308A) at an interface presenting Phe and Tyr residues.

Crystal structure of MCP(W308A) confirms a conserved HK97 fold

The data presented above demonstrate that MCP(W308A) is folded, soluble and retains SP binding interactions. To probe for structural features that may explain the assembly-deficient phenotype, we next determined the high-resolution structure of the mutant protein. The crystal structure was solved in spacegroup C121 at 2.70 Å resolution as described in Materials and Methods (see Table S2 for crystal data and refinement statistics). The asymmetric unit consists of 4 copies of MCP(W308A) (Figure S2A) and while the overall structures of chains A to C are well defined and consistent, there is some conformational heterogeneity displayed in a few loops (Figure 3A). In contrast, chain D has discontinuous electron density with greater RMSD (Figure S2B) and higher B-factors (Figure S2C) than the other three chains. Chain A has the lowest overall B-factors and no rotamer outliers and unless otherwise noted, Chain A was used in all subsequent analyses.

The structure of MCP(W308A) is consistent with our published model²⁷ and shares the conserved HK97 capsid protein fold observed in both prokaryotic and eukaryotic dsDNA viruses.^{31,41} This includes an axial “A-domain” composed of 5 helices surrounding a 6 strand β -sheet (β_4) and a peripheral “P-domain” featuring the prominent “spine helix” cradled by a pair of three-strand β -sheets (β_1 and β_3) (Figures 3B, S3). The residues surrounding the Trp308→Ala substitution exhibit

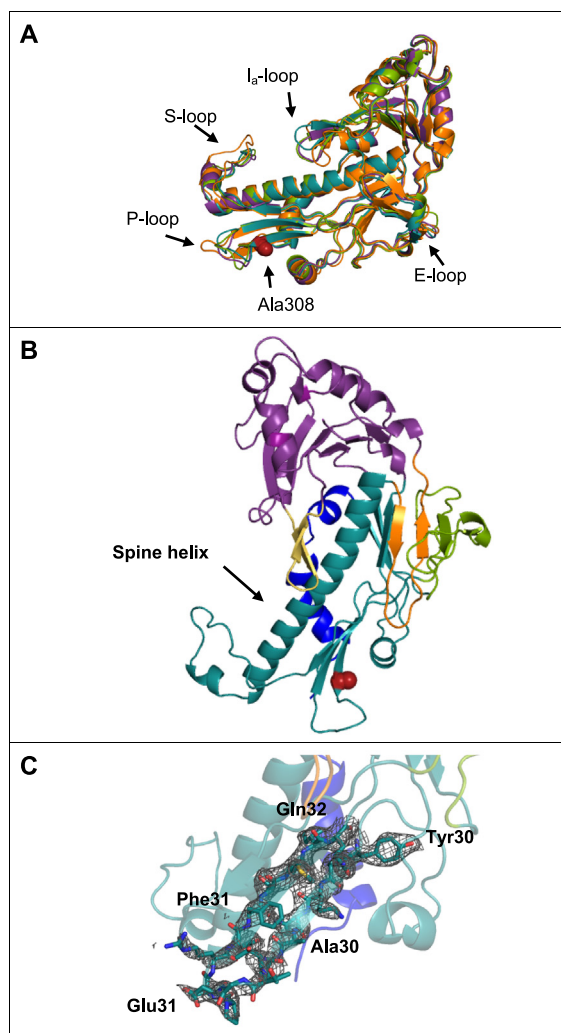


Figure 3. MCP(W308A) Shares the Iconic HK97 Fold. *Panel A.* Superposition of Chains A-D in the MCP (W308A) ASU; chain A, cyan; chain B, green; chain C, magenta; chain D, orange. *Panel B.* MCP(W308A) shares the HK97 fold. The A-domain is colored purple (residues 157–279 and 331–341); the P-domain containing the spine helix is colored teal (residues 75–137 and 280–330). Additional features include^{28,71}; the I_α-loop is colored gold (residues 255–266); the E-loop is colored green (residues 37–74); the G-loop is colored orange (residues 138–156); the N-arm is colored blue (residues 2–28). *Panel C.* Electron density map surrounding Ala308. The 2Fo-Fc map contoured at 1.0 σ is shown in grey. Electron density from Ala308 to Phe318 is poorly defined.

elevated B-factors (Figure S2C) and weaker electron density (Figure 3C) indicating that the area surrounding the mutation is flexible. Indeed, this region is largely surrounded by solvent with few crystal contacts (data not shown) and differs slightly between chains A, B, and C (Figure 3A).

While the overall HK97 fold is conserved, structural differences are observed between

monomeric MCP(W308A) and all other MCP structures that have been determined in the context of a fully assembled icosahedral shell.^{31,42–43} First, MCP(W308A) features a β -hairpin protruding from the A-domain (the I_α loop²⁸ that is not present in any other MCP structures except that of phage YSD1 (compare Figures 3A, 3B and S5F).⁴¹ Second, the conserved “E-loop”, so named because it is extended in the context of an assembled shell (Figure 4), is folded back and interacts with an elongated β -hairpin extending from β 2 in MCP(W308A) (Figure 3); this β -hairpin is homologous to the glycine rich “G loop” in HK97.^{28,44} Third, the N-terminal “N-arm”, is mostly helical and tucked up against the P-domain in MCP(W308A), a conformation also observed in the *procapsid* structures of phages λ , HK97 and P23-45 (Figures 4A, 4B, 4C, respectively).^{45–46} In contrast, the N-arm is extended and interacts with neighboring proteins in the context of a *mature capsid* (Figures 4D–4F).

MCP(W308A) shares a fold similar to monomeric MCPs

The isolated MCPs from phage YSD1 (YSD1_17) and the phage T4 penton protein (gp24) are also monomers in solution, but unlike MCP(W308A), they retain a shell assembly phenotype. To understand whether any gross structural anomalies prevent MCP(W308A) from assembling into capsids, we compare its structure to those of the gp24 and YSD1_17 assembly-competent monomers. The N-arms in all three soluble monomers are mostly helical and closely hugging the P-domains (Figure 5), reminiscent of MCP conformations observed in assembled *procapsid* shells (Figures 4A–4C). Similarly, the E-loops in the soluble MCPs are folded inwards and pressed against the P-domain. In contrast, these compact E-loops are extended in all shell structures (Figure 4) to provide critical inter-subunit interactions required for shell stability.

Within this context, the crystal structure of a “putative capsid protein of prophage (*E.coli* CFT073)” has been reported (PDB ID 3BQW), which shares the compact features of MCP (W308A) (Figure 4D). Indeed, we previously used these data as a basis to model the structure of λ MCP-WT²⁷ and it was used as the search model for molecular replacement in the present study (Materials and Methods). Unfortunately, experimental details pertaining to the CFT073 phage structure have not been published and neither its association state in solution nor its capacity to assemble into shells is known. Based on the available structures described above, we propose that the four monomer structures represent a conserved pre-assembly conformation. That MCP(W308A) does not assemble beyond the dimer state suggests that the W308A mutation traps the protein in this pre-assembly conformation.

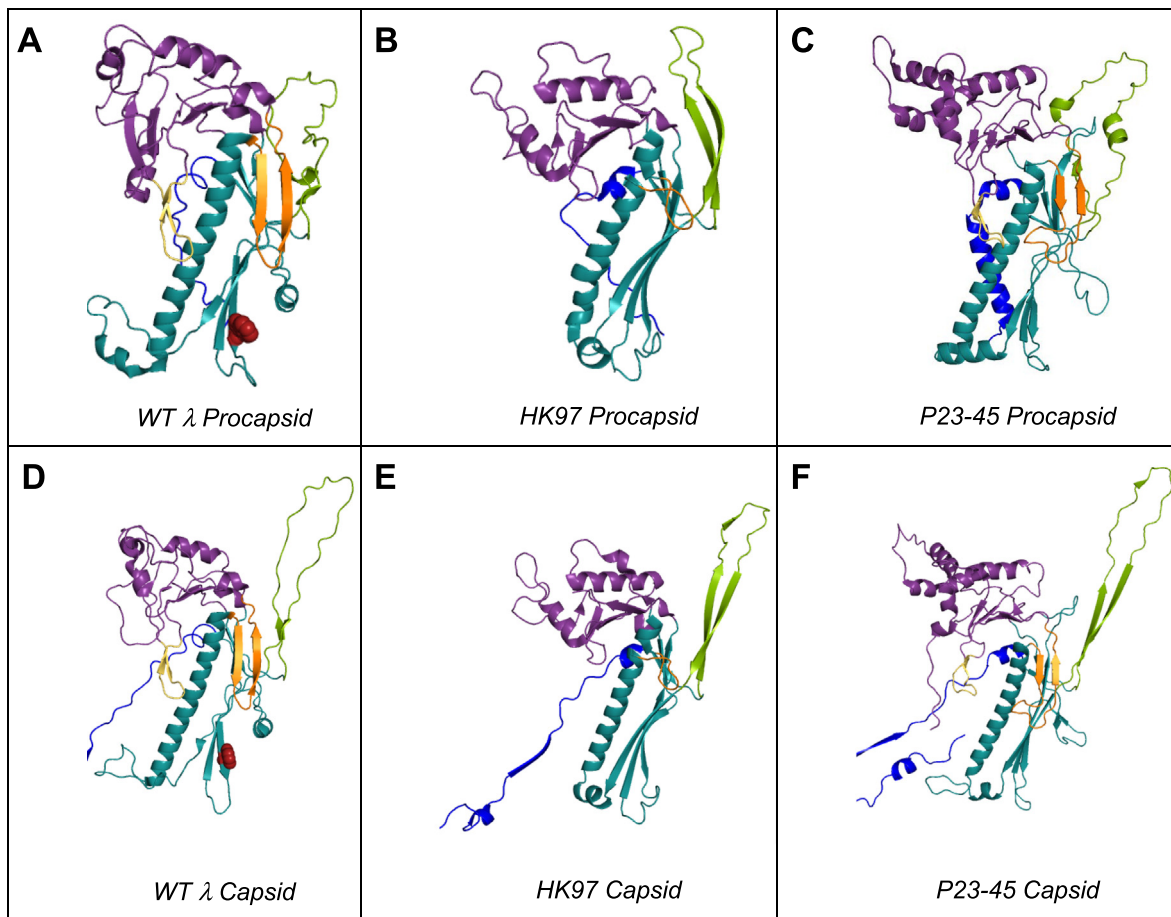


Figure 4. The Soluble MCP(W308A) Monomer Adopts a More Compact Structure Than Observed in Assembled Procapsid and Capsid Shells. *Panel A.* Phage λ wild type MCP monomer (gpE) extracted from the procapsid cryo-EM structure (PDB ID 7VI9). *Panel B.* HK97 MCP monomer (gp5) extracted from the HK97 procapsid structure (PDB ID 3E8K). *Panel C.* P23-45 MCP monomer (gp89) extracted from the procapsid structure (PDB ID 6IBC). *Panel D.* Phage λ wild type MCP monomer (gpE) extracted from the cryo-EM capsid structure (PDB ID 7VII). *Panel E.* MCP monomer (gp5) extracted from the HK97 capsid structure (PDB ID 1OHG). *Panel F.* MCP monomer (gp89) extracted from the P23-45 capsid structure (PDB ID 6I9E). All structural motifs are colored as in [Figure 3B](#).

Structure of λ MCP assembled into a procapsid shell

A high resolution cryo-EM structure of the λ procapsid has recently been published²⁸ and to probe for features that may explain why MCP (W308A) does not assemble into a procapsid shell, we superimposed the MCP(W308A) crystal structure onto the wild-type MCP structure extracted from the recently published cryo-EM reconstruction. The global folds of the soluble MCP(W308A) monomer and wild type λ MCP assembled in the procapsid are remarkably similar ([Figure 6A](#)), which confirms that the mutation does not induce a major structural anomaly that prevents shell assembly. As anticipated, the most striking difference between them is the position of the E-loop, which breaks its interaction with the P-domain in the soluble monomer and swings out in order to interact with a neighboring subunit in the procapsid shell ([Figures 6A, 6B](#)).

Despite its assembly incompetence, MCP (W308A) associates as a tetramer in the crystal asymmetric unit (ASU) in a form reminiscent of a partially assembled capsid hexamer but wherein the E-loops remain in their “tucked in” conformation ([Figure S2A](#)). Superposition of the MCP(W308A) crystal ASU and the cryo-EM ASU of the λ procapsid²⁸ reveals that the two structures overlay quite well, though the subunits in the crystal are more tightly packed and slightly rotated relative to their positions in the shell ([Figure 6C, Table S3](#)). This suggests that while MCP(W308A) retains an inherent native self-association behavior, the E-loops fail to extend to provide critical intersubunit interactions required for further shell assembly.

The Trp308->ala mutation does not alter nanosecond dynamics of MCP

The biophysical and structural data demonstrate that MCP(W308A) is a soluble monomer that

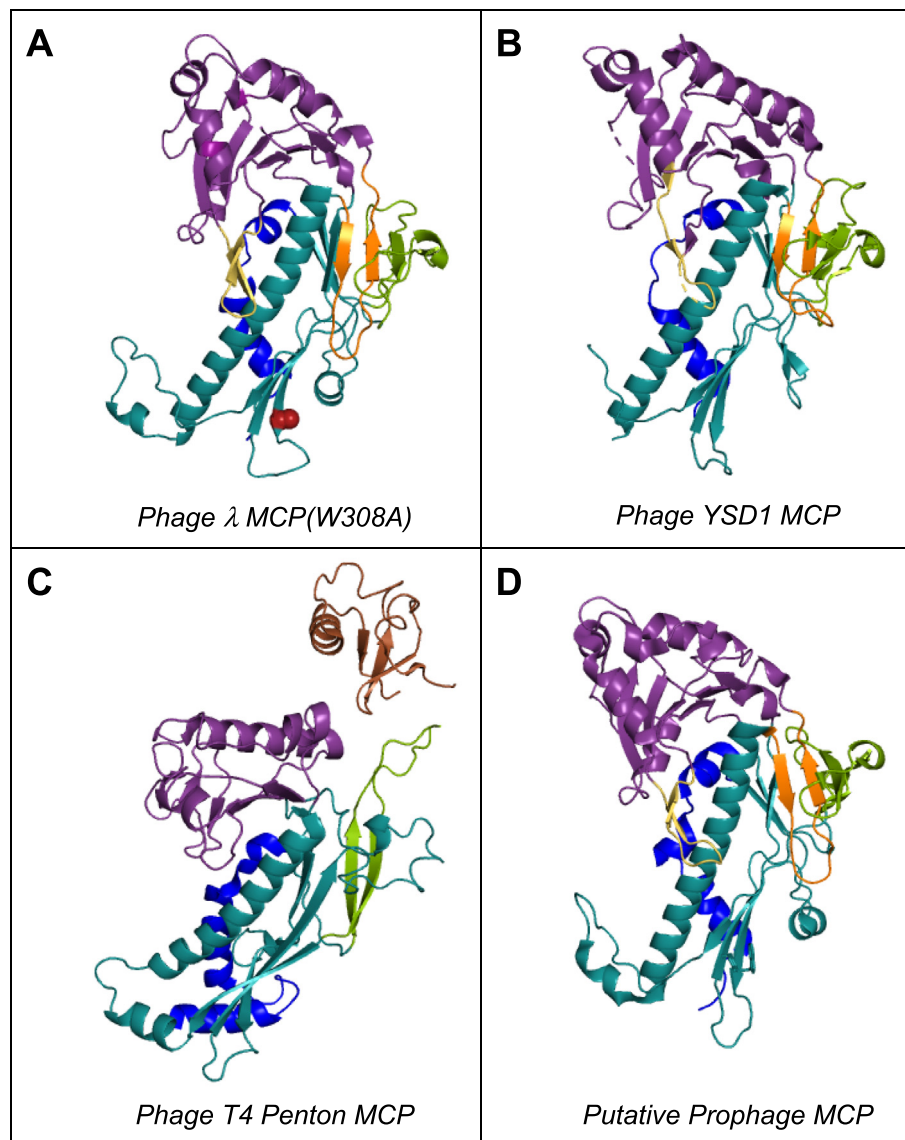


Figure 5. Comparison of MCP(W308A) to Other Soluble Monomeric Capsid Proteins. *Panel A.* MCP(W308A) monomer with Ala308 shown as red spheres. *Panel B.* YSD1_17 MCP monomer from phage YSD1 (PDB ID 6XGP). *Panel C.* gp24 monomer from phage T4 (PDB ID 1YUE). *Panel D.* “Putative capsid protein of prophage (*E.coli* CFT073)” (PDB ID 3BQW). Domains are colored as described in Figure 3B.

retains the conserved HK97 MCP fold. It also appears to maintain the propensity to self-associate in a native, capsomer-like fashion and it retains SP binding interactions, yet the mutant protein fails to assemble beyond the dimer state in solution. Unfortunately, the wild-type protein aggregates and it is not possible to directly compare the two structures to probe for structural features that may explain this assembly deficient phenotype. We therefore employed a computational approach to probe for dynamic effects of the W308A mutation.

We first constructed a wild-type structural model (MCP-WT) by changing Ala308 back to tryptophan *in silico*. The MCP(W308A) and MCP-WT structures then were used as a starting point for

the dynamic simulations as described in Materials and Methods. The data displayed in Figure S6 reveals that protein secondary structures, intra-molecular protein hydrogen bonds and protein-water hydrogen bonds remain intact for both the WT and W308A mutant proteins during the 150 ns simulation period. Moreover, neither the Root Mean Square Deviation (RMSD) nor the Radius of Gyration vary appreciably during the simulation time course. Superposition of the equilibrated (150 ns) WT and W308A models shows that they possess global folds that do not differ significantly from each other (Figure 7A), and that they both retain the global fold observed in the MCP (W308A) crystal structure and MCP-WT assembled into a procapsid shell (Figure S7).

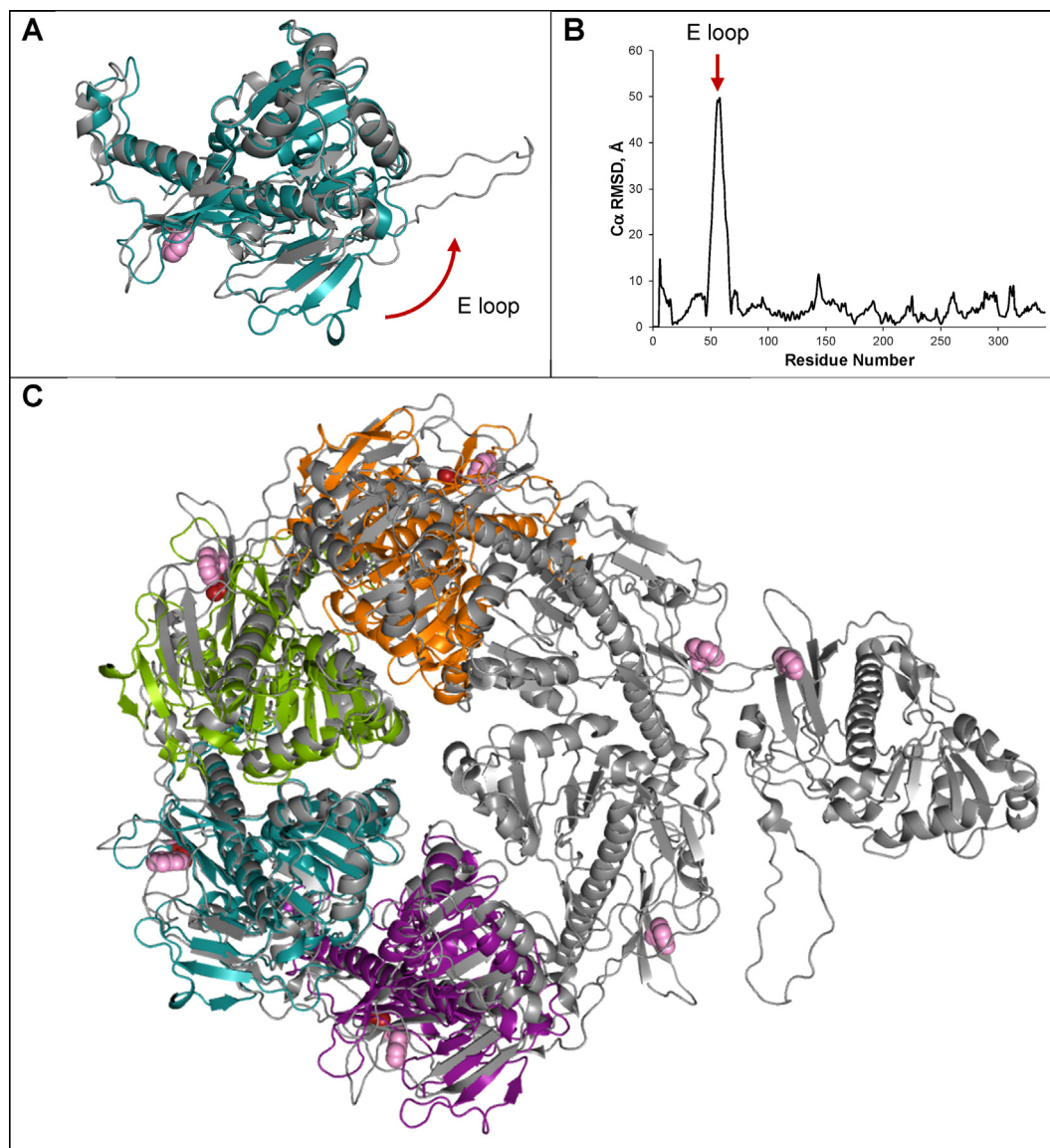


Figure 6. Comparison of Soluble MCP(W308A) and WT-MCP in the Procapsid. *Panel A.* Alignment of MCP (W308A) with WT-MCP demonstrates large-scale change in E-loop conformation upon assembly into a procapsid. *Panel B.* Per residue C α RMSD of aligned structures in Panel A. *Panel C.* Overlay of MCP(W308A) crystal ASU with WT-MCP procapsid ASU (PDB ID 7VI9). Chain A is shown in teal, B in green, C in purple, and D in orange with Ala308 shown as red spheres in the MCP(W308A) crystal ASU. WT-MCP ASU is colored in grey with Trp308 shown as pink spheres. The entire crystal ASU was overlaid on the procapsid cryo-EM ASU by aligning Chain A of the crystal to Chain F of the procapsid using the align command in Pymol.

That said, two regions show modest differences, which are reflected in the Root Mean Square Fluctuation analysis (RMSF) (Figure 7B). First, the P-loop, which is proximate to residue 308, shows a modest shift in both equilibrated models relative to the crystal structure, and this region is more dynamic in the W308A simulation. This is consistent with the crystal structure data that also reveals higher mobility in this loop (Figures 3A, 3C). Second, the S-loop, which precedes the spine helix, similarly shows a shift in both equilibrated models and is *less* dynamic in the W308A simulation relative to wild type (Figure 7).

This is also consistent with the structural data which reveal conformational differences between the chains in the crystallographic ASU (Figure 3A). Given that the W308A mutation little affects the overall structure of the protein, we propose that these dynamic features are related to the assembly defect in the mutant MCP.

Nature of the dynamic defect in MCP(W308)

Close inspection of the MCP(W308A) structure reveals a key intramolecular interaction in the λ MCP – a methionine di-aromatic hydrophobic

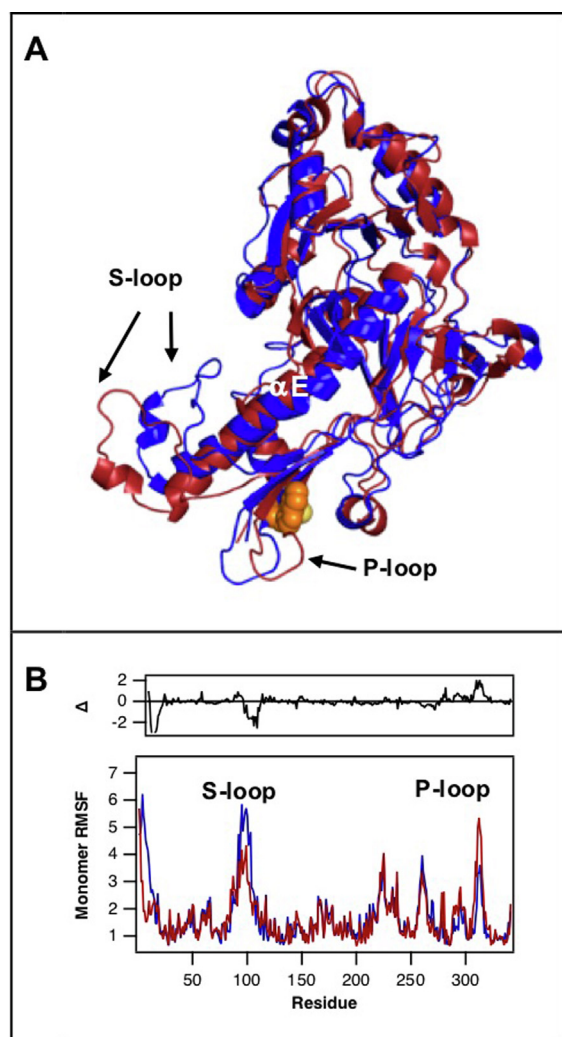


Figure 7. Evaluation of MCP(W308A) and WT Monomers via Molecular Dynamics Simulations. *Panel A.* Final structures of MCP(W308A) and MCP-WT taken from the last frame of the MD simulations (150 ns) are shown in red and blue, respectively. Residue 308 is colored yellow and orange for the mutant and WT models, respectively. *Panel B.* RMSF (indicator of dynamic behavior of the region) versus residue for the MD simulations with data for the MCP(W308A) and WT models shown in red and blue, respectively. Areas of notable difference are labeled with the structural feature affected. Δ shown at top represents MCP(W308A) minus MCP-WT data.

interaction (Met320-Trp308-Phe318) as depicted in Figure 8. Met-aromatic and bridging met-di-aromatic interactions have been identified in approximately one-third of the proteins in the protein data bank and are estimated to provide 1–1.5 kcal/mol folding energy.^{47–48} In the native MCP, this interaction is predicted to “staple” the ends of the P-loop to provide structural rigidity. This stabilizing interaction is compromised by mutation of Trp308 to Ala, which is reflected in the simula-

tions as an increase in the dynamic behavior of the loop (Figure 7) and experimentally as an *increase* in the mobility of the loop in the crystal structure (Figure 3C).

In contrast, the W308A mutation *decreases* the dynamic behavior of the S-loop, which is distant in the primary sequence. Specifically, the mutation lies within strand 15 of the three-strand β 3 sheet (Figure S3). Of note, strand 4 of this sheet (residues 81–83) is positioned directly between the S-loop and the E-loop and we posit that these residues are dynamically coupled), as depicted in Figure 8. Thus, the observed stabilization of the S-loop is coupled to stabilization of the tucked in conformation of the E-loop, thermodynamically trapping the mutant protein in the pre-assembly state. This leads to a failure to sample the conformational states necessary to extend the E-loop, thus halting capsid shell assembly. We note that these large-scale motions likely occur on the microsecond timescale and are not captured on the nanosecond timescale of our current simulations.

Discussion

We have described the biophysical, structural and dynamic characteristics of an assembly incompetent mutant of the bacteriophage λ major capsid protein, MCP(W308A). The structural data reveals that MCP(W308A) generally shares the iconic HK97 fold observed with MCPs assembled into capsid shells, but with significant differences in the N-arm and E-loop residues which are tucked in, tightly interacting with the P-domain. The structure is mirrored in two other soluble, monomeric MCPs suggesting that this represents a conserved pre-assembly conformation; however, while the YSD1 and T4 MCPs retain an assembly competent phenotype, the λ mutant does not assemble beyond the dimer state, even though it retains SP binding interactions. Specifically, the mutation stabilizes the E-loop in the pre-assembly conformation, precluding its extension and abrogating requisite intermolecular interactions required for capsomer assembly.

Cryo-EM studies in phages P22 and 80 α , and with the N-terminal scaffolding delta domain of HK97 reveal that SPs interact with the N-terminal arms of MCPs positioned at the procapsid shell interior.^{19–22} Within this context, the λ MCP N-terminus is rich in Tyr and Phe residues, that are solvent exposed and positioned at the inner surface of the assembled shell (Figure S8). CD-monitored titration data indicate that λ SP binding affects MCP tertiary structures in the primordial assembly complex, ostensibly those in the N-terminus of MCP(W308A), and these critical interactions appear to be retained in the mutant MCP. Thus, neither structural aberrations nor SP binding interactions appear to be significantly affected by the

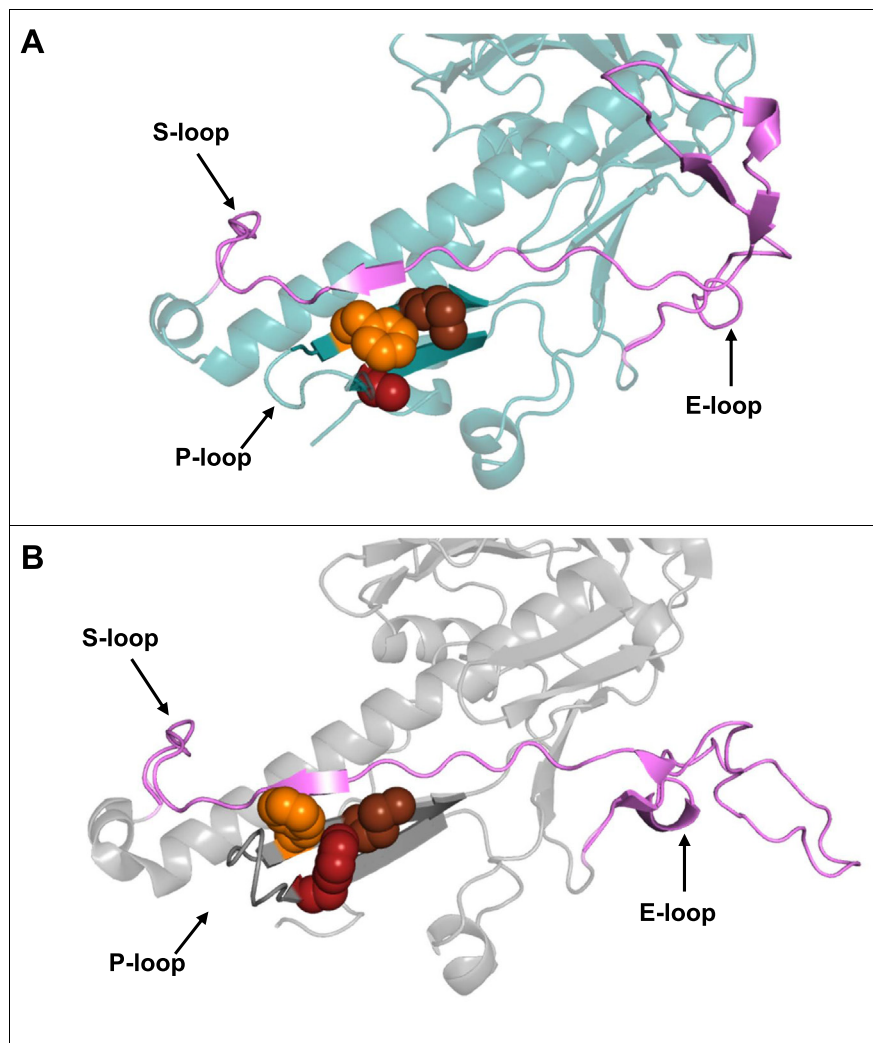


Figure 8. MCP S- and E-loop Residues are Dynamically Coupled. MCP(W308A) and MCP-WT taken from the procapsid structure are shown in Panels A and B, respectively. The W308A mutation lies within strand 15 of the three-strand $\beta 3$ sheet (see **Figure S3**). Strand 4 of this sheet (residues 81–83) is positioned directly between the S-loop and the E-loop. Residues comprising the intramolecular methionine-diaromatic interaction are shown as spheres with Met320 in brown, Phe318 in orange and Ala/Trp308 in red. The Met-diaromatic interaction is disrupted with the Try->Ala mutation, resulting in movement of Phe318 to establish a weaker, Met-aromatic interaction in the mutant protein.

W308A mutation and the ensemble of data suggest that the assembly defect is a dynamic one.

The assembly incompetence of MCP(W308A) provides insight into primordial intermediates leading to MCP-SP co-polymerization and shell assembly, and we propose the model depicted in **Figure 9**. The critical intermediate is a heterotrimer composed of one SP dimer bound to one MCP monomer ($\text{SP}_{\text{dimer}} \bullet \text{MCP}$), as characterized above. While we have defined several thermodynamic binding constants and determined a macroscopic $K_{D, \text{app}} \sim 26 \mu\text{M}$ for SP and MCP assembly, the microscopic binding constants cannot be determined from the present data. Nevertheless, the likely pathway for $\text{SP}_{\text{dimer}} \bullet \text{MCP}$ assembly is depicted in dark arrows. Subsequent co-polymerization of $\text{SP}_{\text{dimer}} \bullet \text{MCP}$ and/or $\text{SP}_{\text{dimer}} \bullet \text{MCP}_2$ intermediates ultimately affords

capsomers that make up the icosahedral shell. Extension of the E-loop from its tucked-in position to an extended conformation that interacts with a neighboring MCP subunit is critical for the stability of the assembly intermediates and is required for capsomer assembly. That this conformational switch does not occur with MCP(W308A) results in a reversible interaction that fails to progress beyond the primordial $\text{SP}_{\text{dimer}} \bullet \text{MCP}$ intermediate, thus abrogating shell assembly the earliest steps.

Virus capsid assembly proceeds through a reversible, rate limiting nucleation step followed by rapid polymerization of MCP proteins.^{49–50} In some cases, such as hepatitis B, shell polymerization proceeds through a scaffolding protein independent polymerization of MCP dimers⁵¹. In other cases, such as the herpesviruses and phage P22, the scaffolding proteins copolymerize with capsid protein

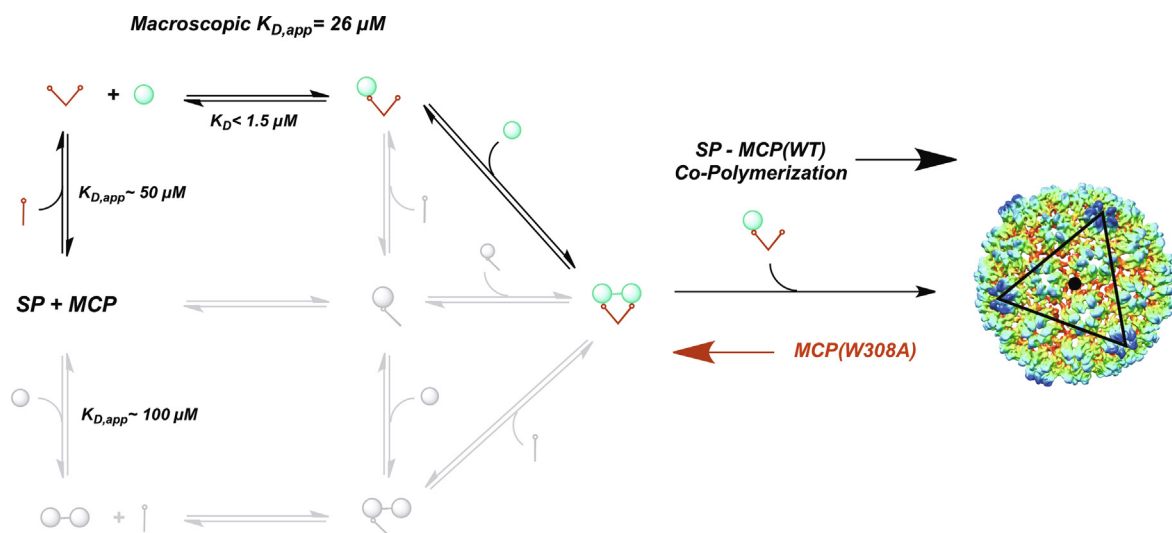


Figure 9. Proposed Model for Shell Assembly by Scaffold and Major Capsid Proteins. The $SP_{dimer} \bullet MCP$ intermediate represents the primordial shell initiation species. Experimentally defined binding constants are indicated. Dark black arrows depict the preferred assembly pathway with a macroscopic binding constant for SP and MCP assembly into the $SP_{dimer} \bullet MCP$ intermediate, $K_{D,app} = 26 \pm 2 \mu M$. Details provided in text.

monomers that cooperatively assemble into the shells.^{31,37,51} We propose that the phage λ shell grows by addition of MCP monomers, chaperoned by SP dimers. In all cases, high-fidelity procapsid assembly is proposed to occur via multiple weak interactions between the MCP proteins adding to the nascent shell. Thus, shell assembly may be considered an equilibrium polymerization reaction^{16,52}, which allows for reversible “sampling” of the shell assembly intermediates to prevent off-pathway, dead-end shell products.⁵³ As such, relatively minor insults to MCP structure could affect essential dynamic interactions and have major effects on the polymerization reactions, especially if they affect the earliest assembly steps⁵² and as shown here for the MCP(W308A) mutant MCP.

The present data clearly indicate the MCP (W308A) mutation abrogates shell assembly and that the $SP_{dimer} \bullet MCP$ is an early and essential intermediate in the shell polymerization reactions. Given that the essential features of scaffold chaperoned MCP polymerization are strongly conserved amongst all of the complex dsDNA viruses, from phages to herpesviruses, these results have broad biological significance in our understanding of virus development.

Materials and Methods

Detailed protocols for the purification of proteins used in this study are described in the Supplemental Materials. Absorbance spectra were obtained using a Thermo Scientific NanoDrop UV-vis spectrophotometer (San Jose, CA, USA). CD spectra were recorded on a Chirascan V100 circular dichroism spectrometer (Applied Photophysics; Leatherhead, Surrey, UK).

Analytical ultracentrifugation studies were performed in a Beckman Coulter Optima X-LA analytical ultracentrifuge. Protein purifications utilized an ÄKTA Purifier chromatography system (GE Healthcare). HiTrapQ HP, HiTrapSP and Superose 6 Increase columns were purchased from GE Healthcare. Amicon® centrifugal filters were purchased from Sigma-Aldrich.

Crystallization of MCP(W308A)

Purified MCP(W308) in storage buffer was exchanged into Buffer A (10 mM Tris buffer, pH 8 at 4 °C, containing 20 mM NaCl, 10 mM MgCl₂, 1 mM EDTA and 1 mM TCEP) by size exclusion chromatography (SEC). Protein samples (500 mL) were loaded onto a Superdex S-200 10/300 GL column equilibrated and eluted with Buffer A. Elution fractions were examined by SDS-PAGE and those containing MCP(W308A) were pooled and concentrated to ~ 300 μM using 10 kDa MWCO Amicon Centrifugal Filters (Millipore Sigma) at 4 °C according to manufacturer's directions. The purified protein was used to set up a JCSG + crystallization screen (Molecular Dimensions) in 96 well plates (Axygem) using a Phoenix drop-setter robot at 18 °C. The samples were incubated at 18 °C and after 3 days microcrystals formed in 0.1 M Bis-tris buffer, pH 5.5, containing 0.1 M ammonium sulfate and 25 % w/v PEG 3350. Crystallization conditions were optimized using hanging drop plates that were prepared using 0.1 M bis-tris buffer, pH 5.7, containing 0.1 M ammonium sulfate, 5 % v/v glycerol, and 30 % v/v PEG 3350. The hanging drops were streak seeded with the microcrystals from above and incubated at 18 °C. Rhomboid

prism crystals between 50 and 100 μm formed after 5 days and were mounted and preserved in liquid nitrogen.

MCP(W308A) structure determination

The crystals formed in the space group C121 and data were collected at the Molecular Biology Consortium beamline 4.2.2 at the Advanced Light Source at Lawrence Berkeley National Laboratory. The data were processed using XDS to a resolution 2.70 Å (Table S1). Molecular replacement was performed using PHASER⁵⁴ implemented in Phenix⁵⁵ with the putative capsid protein of a prophage (PDB ID: 3BQW) as the search model. Iterative refinement and model building using Phenix.refine and Coot⁵⁶ gave a final model with good geometry, stereochemistry and structural characteristics as determined by PDB validation analyses⁵⁷ (Table S1). The final structure has PDB ID: 7SJ5. PyMol⁶⁰ was used to measure RMSD values and to prepare figures.

Self-association of MCP(W308A)

Protein samples were dialyzed against 50 mM Tris buffer, pH 8 at 4°C, containing 100 mM NaCl, 20 mM MgCl_2 and 1 mM TCEP and then concentrated using 10 kDa MWCO Amicon Centrifugal Filters according to manufacturer's directions. The samples (410 μl) were then loaded into two sector epon-charcoal cells along with a buffer blank in the reference sector and allowed to equilibrate to 4°C; the protein concentration was as indicated in each individual experiment. The samples were spun at 50,000 rpm and sedimentation was monitored by absorbance at 280 nm. The sedimentation velocity analytical ultracentrifugation (SV-AUC) data were analyzed using the Sedfit program.⁶¹

AUC-monitored SP-MCP(W308A) binding interactions

Protein samples were prepared as described above and mixed to afford binding reaction mixtures containing 15 μM MCP(W308A) and the indicated concentration of SP. The samples were loaded into two sector epon-charcoal cells and SV-AUC was performed as above. The SV-AUC data were analyzed according to the model-independent $g(s)$ approach using the DCDT + program.⁶² The binding data were fit to a phenomenological Hill-Langmuir model (equation (1)) or a stoichiometric binding model (equation (2)), as indicated.

Hill-Langmuir binding model⁶³:

$$s^*(20, w) = \frac{S_{\text{max}} * [\text{SP}]^n}{K_{D, \text{app}} * [\text{SP}]^n} \quad (1)$$

where [SP] is the total concentration of SP added to the binding reaction, S_{max} is $s^*(20, w)$ at saturation, $K_{D, \text{app}}$ is the derived binding constant and n is the Hill coefficient, which was held constant at 1.

Stoichiometric binding model⁶⁴:

$$s^*(20, w) = \frac{(R + K_{D, \text{app}} + n) - \sqrt{(R + K_{D, \text{app}} + n)^2 - 4 * R * n}}{2 * n} \quad (2)$$

where R is the molar ratio of protomer to $\text{SP}_{\text{dimer}} \cdot \text{MCP}$, n is the stoichiometric equivalence point and $K_{D, \text{app}}$ is the apparent binding constant.

CD-Monitored SP-MCP(W308A) Binding Interactions. A 500 μL sample of 30 μM MCP (W308A) prepared as described above was loaded into a 10 mm quartz cuvette (Starna Cells; Atascadero, CA, USA). A solution of 1.5 mM SP was incrementally added to yield the indicated concentration of SP and the samples were allowed to equilibrate at 4°C for at least 2 hours. The spectra from 260–320 nm were averaged from 8 separate scans, were adjusted for concentration and background corrected.

Molecular dynamics simulations

The dynamics studies were conducted using YASARA Structure 19.7 (YASARA Biosciences GmbH, Vienna, Austria; www.yasara.org). Proteins or protein complexes were placed in a simulation cell under periodic boundary conditions, filled with water, 0.9 % NaCl, and additional counter ions to neutralize the system, with a pH 7.4, and at a temperature of 298 K.⁶⁵ The main MD simulation was run for 75 or 150 nanoseconds, as indicated, using the Amber (ff14SB) force field⁶⁶ with GAFF⁶⁷/AM1BCC⁶⁸ parameters, particle mesh Ewald summation, an 8.0-Å cutoff for nonbonded forces, a 5-femtosecond time step, and LINCS-constrained hydrogen atoms⁶⁹ and kept at constant pressure and temperature (the NPT ensemble), as described previously.⁷⁰

Accession Number

The coordinates of MCP(W308A) have been deposited in the Protein Data Bank with accession code 7SJ5.

CRedit authorship contribution statement

Christal R. Davis: Conceptualization, Methodology, Formal analysis, Investigation, Visualization, Writing – original draft, Writing – review & editing, Funding acquisition. **Donald Backos:** Methodology, Software, Resources. **Marc C. Morais:** Supervision, Writing – review & editing. **Mair E.A. Churchill:** Formal analysis,

Supervision, Writing – review & editing. **Carlos E. Catalano:** Conceptualization, Methodology, Resources, Visualization, Supervision, Writing – original draft, Writing – review & editing, Funding acquisition.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors wish to express their appreciation to Jay Nix (Beamline 4.2.2 at the Advanced Light Source, Lawrence Berkeley National Laboratory) and John Hardin (Structural Biology Shared Resource, University of Colorado Cancer Center) for data collection and technical support of the x-ray structural studies.

Funding

This work was supported by National Science Foundation grants MCB 2016019 (CEC) and GRFP DGE-1938058 (CRD); National Institutes of Health grant R01GM127365 (CEC and MCM); and by the National Institutes of Health grants P30CA046934 (Cancer Center Support Grant) and S10OD012033 (UC Cancer Center).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jmb.2022.167719>.

Received 26 February 2022;

Accepted 6 July 2022;

Available online 9 July 2022

Keywords:

virus capsid assembly;
major capsid protein;
shell nucleation complexes;
HK97 fold;
scaffolding protein

References

1. Virus, B.A., Strategies, R., (2007). In: *Fields Virology*. Lippincott-Williams, and Wilkins, New York, pp. 119–138.
2. Catalano, C.E., (2005). Viral Genome Packaging Machines: An Overview. In: Catalano, C.E. (Ed.), *Viral Genome Packaging Machines: Genetics, Structure, and Mechanism*. New York, NY, Springer, Boston, MA, pp. 1–4.
3. Casjens, S.R., (2011). The DNA-packaging nanomotor of tailed bacteriophages. *Nature Rev Micro* **9**, 647–657.
4. Rao, V.B., Feiss, M., (2015). Mechanisms of DNA Packaging by Large Double-Stranded DNA Viruses. *Ann Rev Virol* **2**, 351–378.
5. Casjens, S.R., Hendrix, R.W., (2015). Bacteriophage lambda: early pioneer and still relevant. *Virology*, 310–330.
6. Roizman, B., Knipe, D.M., Whitley, R.J., (2007). Herpes Simplex Viruses. In: Knipe, D.M., Howley, P.M. (Eds.), *Fields Virology*. Fifth ed. New York, NY, Lippincott, Williams, and Wilkins, pp. 2501–2602.
7. Myers, M.W., Carter, B.J., (1980). Assembly of Adeno-Associated Virus. *Virology* **102**, 71–82.
8. Murialdo, H., Siminovitch, L., (1972). The Morphogenesis of Phage Lambda. V. Form-Determining Function of the Genes Required for the Assembly of the Head. *Virology* **48**, 824–835.
9. Ray, P., Murialdo, H., (1975). The role of gene Nu3 in bacteriophage lambda head morphogenesis. *Virology* **64**, 247–263.
10. Prevelige, P.E., Fane, B.A., (2012). Building the Machines: Scaffolding Protein Functions During Bacteriophage Morphogenesis. In: Rossmann, M.G., Rao, V.B. (Eds.), *Viral Molecular Machines*. Springer, US, Boston, MA, pp. 325–350.
11. Zlotnick, A., Fane, B.A., (2011). Mechanisms of Icosahedral Virus Assembly. In: Agbandje-McKenna, M., McKenna, R. (Eds.), *Structural Virology*. The Royal Society of Chemistry, Cambridge, UK, pp. 182–204.
12. Fane, B.A., Prevelige, P.E., (2003). Mechanism of Scaffolding-Assisted Viral Assembly. In: Wah, C., John, E.J. (Eds.), *Virus Structure*. Academic Press, Oxford, UK, pp. 259–299.
13. Bamford, D.H., Grimes, J.M., Stuart, D.I., (2005). What does structure tell us about virus evolution? *Curr Opin Struct Biol* **15**, 655–663.
14. Steven, A.C., Heymann, J.B., Cheng, N., Trus, B.L., Conway, J.F., (2005). Virus Maturation: Dynamics and Mechanism of a Stabilizing Structural Transition that Leads to Infectivity. *Curr Opin Struct Biol* **15**, 227–236.
15. Calendar, R., Abedon, S.T., (2006). The Bacteriophages. Oxford University Press, New York, N.Y..
16. Johnson, J.E., (2010). Virus Particle Maturation: Insights into Elegantly Programmed Nanomachines. *Curr Opin Struct Biol* **20**, 210–216.
17. Catalano, C.E., (2018). Bacteriophage lambda: The path from biology to theranostic agent. *Wiley Interdiscip Rev Nanomed Nanobiotechnol* **10**, e1517.
18. Duda, R.L., Teschke, C.M., (2019). The amazing HK97 fold: versatile results of modest differences. *Curr Opin Virol* **36**, 9–16.
19. Dearborn, A.D., Wall, E.A., Kizziah, J.L., Klenow, L., Parker, L.K., Manning, K.A., (2017). Competing scaffolding proteins determine capsid size during mobilization of *Staphylococcus aureus* pathogenicity islands. *Elife* **6**.
20. Chen, D.-H., Baker, M.L., Hryc, C.F., DiMaio, F., Jakana, J., Wu, W., (2011). Structural Basis for Scaffolding-Mediated Assembly and Maturation of a dsDNA Virus. *Proc Natl Acad Sci* **108**, 1355–1360.
21. Morais, M.C., Kanamaru, S., Badasso, M.O., Koti, J.S., Owen, B.A., McMurray, C.T., (2003). Bacteriophage phi29 Scaffolding Protein gp7 Before and After Prohead Assembly. *Nat Struct Biol* **10**, 572–576.

22. Huang, R.K., Khayat, R., Lee, K.K., Gertsman, I., Duda, R. L., Hendrix, R.W., (2011). The Prohead-I Structure of Bacteriophage HK97: Implications for Scaffold-Mediated Control of Particle Assembly and Maturation. *J Mol Biol* **408**, 541–554.
23. Hendrix RW, Roberts JW, Stahl FW, Weisberg RA. Lambda II. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory; 1983.
24. Murialdo, H., Becker, A., (1978). Head Morphogenesis of Complex Double-Stranded Deoxyribonucleic Acid Bacteriophages. *Microbiol Rev* **42**, 529–576.
25. Georgopoulos, C., Tilly, K., Casjens, S., (1983). Lambdoid Phage Head Assembly. In: Hendrix, R.W., Roberts, J.W., Stahl, F.W., Weisberg, R.A. (Eds.), *Lambda II*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, pp. 279–304.
26. Medina, E., Wiczeorek, D.J., Medina, E.M., Yang, Q., Feiss, M., Catalano, C.E., (2010). Assembly and Maturation of the Bacteriophage Lambda Procapsid: gpC Is the Viral Protease. *J Mol Biol* **401**, 813–830.
27. Singh, P., Nakatani, E., Goodlett, D.R., Catalano, C.E., (2013). A Pseudo-Atomic Model for the Capsid Shell of Bacteriophage Lambda Using Chemical Cross-Linking/ Mass Spectrometry and Molecular Modeling. *J Mol Biol* **425**, 3378–3388.
28. Wang C, Zeng J, Wang J. Structural basis of bacteriophage lambda capsid maturation. *Structure* 2022;30:637-45.e3.
29. Medina, M.M., Andrews, B.T., Nakatani, E., Catalano, C. E., (2011). The Bacteriophage Lambda gpNu3 Scaffolding Protein is an Intrinsically Disordered and Biologically Functional Procapsid Assembly Catalyst. *J Mol Biol* **412**, 723–736.
30. Lambert, S., Yang, Q., De Angeles, R., Chang, J.R., Ortega, M., Davis, C., (2017). Molecular Dissection of the Forces Responsible for Viral Capsid Assembly and Stabilization by Decoration Proteins. *Biochemistry* **56**, 767–778.
31. Suhanovsky, M.M., Teschke, C.M., (2015). Nature's favorite building block: Deciphering folding and capsid assembly of proteins with the HK97-fold. *Virology* **479–480**, 487–497.
32. Yang, D., Correia, J.J., Stafford Iii, W.F., Roberts, C.J., Singh, S., Hayes, D., (2018). Weak IgG self- and hetero-association characterized by fluorescence analytical ultracentrifugation. *Protein Sci* **27**, 1334–1348.
33. Cole, J.L., Correia, J.J., Stafford III, W.F., (2011). *The Use of Analytical Sedimentation Velocity to Extract Thermodynamic Biophysical Chemistry* **159**, 120–128.
34. Gaussier, H., Yang, Q., Catalano, C.E., (2006). Building a virus from scratch: assembly of an infectious virus using purified components in a rigorously defined biochemical assay system. *J Mol Biol* **357**, 1154–1166.
35. Parker, M.H., Stafford, W.F., Prevelige, P.E., (1997). Bacteriophage P22 Scaffolding Protein forms Oligomers in Solution. *J Mol Biol* **268**, 655–665.
36. Poh, S.L., el Khadali, F., Berrier, C., Lurz, R., Melki, R., Tavares, P., (2008). Oligomerization of the SPP1 Scaffold Protein. *J Mol Biol* **378**, 551–564.
37. Newcomb, W.W., Homa, F.L., Thomsen, D.R., Trus, B.L., Cheng, N., Steven, A., (1999). Assembly of the Herpes Simplex Virus Procapsid from Purified Components and Identification of Small Complexes Containing the Major Capsid and Scaffolding Proteins. *J Virol* **73**, 4239–4250.
38. Schmidt, U., Darke, P.L., (1997). Dimerization and Activation of the Herpes Simplex Virus Type 1 Protease. *J Biol Chem* **272**, 7732–7735.
39. Darke, P.L., Cole, J.L., Waxman, L., Hall, D.L., Sardana, M.K., Kuo, L.C., (1996). Active Human Cytomegalovirus Protease Is a Dimer. *J Biol Chem* **271**, 7445–7449.
40. Senear, D.F., Dalma-Weiszhausz, D.D., Brenowitz, M., (1993). Effects of anomalous migration and DNA to protein ratios on resolution of equilibrium constants from gel mobility-shift assays. *Electrophoresis* **14**, 704–712.
41. Hardy, J.M., Dunstan, R.A., Grinter, R., Belousoff, M.J., Wang, J., Pickard, D., (2020). The architecture and stabilisation of flagellotropic tailed bacteriophages. *Nat Commun* **11**, 3748.
42. Huet, A., Duda, R.L., Boulanger, P., Conway, J.F., (2019). Capsid expansion of bacteriophage T5 revealed by high resolution cryoelectron microscopy. *Proc Natl Acad Sci* **116**, 21037.
43. Wikoff, W.R., Liljas, L., Duda, R.L., Tsuruta, H., Hendrix, R. W., Johnson, J.E., (2000). Topologically Linked Protein Rings in the Bacteriophage HK97 Capsid. *Science* **289**, 2129–2133.
44. Tso, D.J., Hendrix, R.W., Duda, R.L., (2014). Transient contacts on the exterior of the HK97 procapsid that are essential for capsid assembly. *J Mol Biol* **426**, 2112–2129.
45. Conway, J.F., Wikoff, W.R., Cheng, N., Duda, R.L., Hendrix, R.W., Johnson, J.E., (2001). Virus Maturation Involving Large Subunit Rotations and Local Refolding. *Science* **292**, 744–748.
46. Gertsman, I., Gan, L., Guttman, M., Lee, K., Speir, J.A., Duda, R.L., (2009). An Unexpected Twist in Viral Capsid Maturation. *Nature* **458**, 646–650.
47. Valley, C.C., Cembran, A., Perlmutter, J.D., Lewis, A.K., Labello, N.P., Gao, J., (2012). The methionine-aromatic motif plays a unique role in stabilizing protein structure. *J Biol Chem* **287**, 34979–34991.
48. Weber, D.S., Warren, J.J., (2019). The interaction between methionine and two aromatic amino acids is an abundant and multifunctional motif in proteins. *Arch Biochem Biophys* **672**, 108053.
49. Asor, R., Schlicksup, C.J., Zhao, Z., Zlotnick, A., Raviv, U., (2020). Rapidly Forming Early Intermediate Structures Dictate the Pathway of Capsid Assembly. *J Am Chem Soc* **142**, 7868–7882.
50. Buzón, P., Maity, S., Christodoulis, P., Wiertsema, M.J., Dunkelbarger, S., Kim, C., (2021). Virus self-assembly proceeds through contact-rich energy minima. *Sci Adv* **7** eabg0811-eabg.
51. Zhao, Z., Wang, J.-C.-Y., Zhang, M., Lyktey, N.A., Jarrold, M.F., Jacobson, S.C., (2021;12:589-). Asymmetrizing an icosahedral virus capsid by hierarchical assembly of subunits with designed asymmetry. *Nat Commun*.
52. Katen, S., Zlotnick, A., (2009). The thermodynamics of virus capsid assembly. *Methods Enzymol* **455**, 395–417.
53. Kant, R., Llauró, A., Rayaprolu, V., Qazi, S., de Pablo, P.J., Douglas, T., (2018). Changes in the stability and biomechanics of P22 bacteriophage capsid during maturation. *Biochim Biophys Acta Gen Subj* **1862**, 1492–1504.
54. McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J., (2007). Phaser crystallographic software. *J Appl Crystallogr* **40**, 658–674.

55. Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213–221.
56. Emsley, P., Cowtan, K., (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126–2132.
57. Read, R.J., Adams, P.D., Arendall 3rd, W.B., Brunger, A. T., Emsley, P., Joosten, R.P., (2011). A new generation of crystallographic validation tools for the protein data bank. *Structure* **19**, 1395–1412.
60. DeLano WL. The PyMOL Molecular Graphics System. In: Scientific D, editor. San Carlos, CA2002.
61. Schuck, P., (2000). Size-Distribution Analysis of Macromolecules by Sedimentation Velocity Ultracentrifugation and Lamm Equation Modeling. *Biophys J* **78**, 1606–1619.
62. Philo, J.S., (2006). Improved Methods for Fitting Sedimentation Coefficient Distributions Derived by Time-Derivative Techniques. *Anal Biochem* **354**, 238–246.
63. Michel, D., (2007). Cooperative Equilibrium Curves Generated by Ordered Ligand Binding to Multi-Site Molecules. *Biophys Chem* **129**, 284–288.
64. Rambo, R.P., Doudna, J.A., (2004). Assembly of an Active Group II Intron–Maturase Complex by Protein Dimerization. *Biochemistry* **43**, 6486–6497.
65. Krieger, E., Darden, T., Nabuurs, S.B., Finkelstein, A., Vriend, G., (2004). Making Optimal Use of Empirical Energy Functions: Force-Field Parameterization in Crystal Space. *Proteins* **57**, 678–683.
66. Maier, J.A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K.E., Simmerling, C., (2015). ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J Chem Theory Comput* **11**, 3696–3713.
67. Wang, J., Wolf, R.M., Caldwell, J.W., Kollman, P.A., Case, D.A., (2004). Development and Testing of a General Amber Force Field. *J Comput Chem* **25**, 1157–1174.
68. Jakalian, A., Jack, D.B., Fast, B.C., (2002). Efficient Generation of High-Quality Atomic Charges. AM1-BCC model: II. Parameterization and Validation. *J Comput Chem* **23**, 1623–1641.
69. Hess, B., Bekker, H., Berendsen, H.J.C., Fraaije, J.G.E.M., (1998). LINCS: A linear Constraint Solver for Molecular Simulations. *J Comput Chem* **18**, 1463–1472.
70. Krieger, E., Vriend, G., (2015). New ways to boost molecular dynamics simulations. *J Comput Chem* **36**, 996–1007.
71. Bayfield, O.W., Klimuk, E., Winkler, D.C., Hesketh, E.L., Chechik, M., Cheng, N., (2019). Cryo-EM structure and in vitro DNA packaging of a thermophilic virus with supersized T=7 capsids. *Proc Natl Acad Sci* **116**, 3556.

Further reading

58. Collaborative Computational Project N, (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**, 760–763.
59. Jo, S., Kim, T., Iyer, V.G., Im, W., (2008). CHARMM-GUI: a web-based graphical user interface for CHARMM. *J Comput Chem* **29**, 1859–1865.