Published in final edited form as:

J Am Chem Soc. 2009 February 25; 131(7): 2606–2614. doi:10.1021/ja807730x.

Invariant Polymorphism In Virus Capsid Assembly

Hung D. Nguyen[†], Vijay S. Reddy[†], and Charles L. Brooks III^{*,†}

Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA, Department of Chemistry, Biophysics Program, 930 N. University Ave, University of Michigan, Ann Arbor, MI 48109, USA, and Department of Chemistry and Biophysics Program, 930 N. University Ave, University of Michigan, Ann Arbor, MI 48109, USA

Abstract

Directed self-assembly of designed viral capsids holds significant potential for applications in materials science and engineering. However, the complexity of preparing these systems for assembly and the difficulty of quantitative experimental measurements on the assembly process has limited access to critical mechanistic questions that dictate the final product yields and isomorphous forms. Molecular simulations provide a means of elucidating self-assembly of viral proteins into icosahedral capsids and are the focus of the present study. Using geometrically realistic coarse-grained models with specialized molecular dynamics methods, we delineate conditions of temperature and coat protein concentration that lead to the spontaneous self-assembly of T=1 and T=3 icosahedral capsids. In addition to the primary product of icosahedral capsids, we observe a ubiquitous presence of non-icosahedral yet highly symmetric and enclosed aberrant capsules in both T=1 and T=3 systems. This polymorphism in assembly products recapitulates the scope and morphology of particle types that have been observed in mis-assembly experiments of virus capsids. Moreover, we find that this structural polymorphism in the endpoint structures is an inherent property of the coat proteins and arises from condition-dependent kinetic mechanisms that are independent of the elemental mechanisms of capsid growth (as long as the building blocks of the coat proteins are all either monomeric, dimeric or trimeric) and the capsid T number. The kinetic mechanisms responsible for self-assembly of icosahedral capsids and aberrant capsules are deciphered; the self-assembly of icosahedral capsids requires a high level of assembly fidelity whereas self-assembly of non-icosahedral capsules is a consequence of an off-pathway mechanism that is prevalent under non-optimal conditions of temperature or protein concentration during assembly. The latter case involves kinetically trapped dislocations of pentamer-templated proteins with hexameric organization. These findings provide insights into the complex processes that govern viral capsid assembly suggest some features of the assembly process that can be exploited to control the assembly of icosahedral capsids and non-icosahedral capsules.

Introduction

The spontaneous self-assembly of identical protein units into complex but highly regular icosahedral capsid shells,^{1,2} which protect the packaged viral genome, has motivated studies in virology for decades. More recently, viral capsids have been considered the ultimate biologically ordered system at the nanoscale level, with beneficial applications emerging in

brookscl@umich.edu.

[†]The Scripps Research Institute

[‡]University of Michigan

[§]University of Michigan

Supporting Information Available

Four movies showing oblate, angular, twisted and tubular capsules from T=3 simulations; additional analysis on icosahedral capsids, structural stability and alternative assembly conditions. This material is available free of charge via the Internet at <http://pubs.acs.org>.

materials science and medicine.³ For example, empty viral capsids of the human papilloma virus serve as cervical cancer vaccines,⁴ where potency depends strongly upon the degree of capsid self-assembly.⁵ The inability to control assembly in laboratory and manufacturing procedures lead to architectural contaminants (i.e., structural polymorphism).⁶ A clear understanding of the kinetic mechanisms and thermodynamic control of capsid self-assembly would provide invaluable insights into controlled self-assembly of icosahedral capsids and serve as a crucial prerequisite to their widespread application in medicine and bio-nanotechnology.

In recent years, substantial progress has been made via a plethora of experimental, modeling and simulation-based studies^{7–20} in regards to capsid self-assembly in general. Invaluable kinetic insights have been obtained from several studies^{8,16–18} that examine the formation of icosahedral capsids which are formed with a high degree of absolute assembly fidelity. However, there are only a few studies that focus on the self-assembly of non-icosahedral capsules^{21–23} and the kinetic mechanism for the formation of non-icosahedral capsules that co-exist with icosahedral capsids (e.g., different oblong flock house virus particles observed by Dong *et al.*²⁴) remains unknown. The present computational study was designed to augment this knowledge with the aim of optimizing experimental designs.

It is well recognized that the coat protein shape plays an important role in the organization of the closed icosahedral structure. Statistically, coat proteins largely comprise a conserved β -barrel motif with a trapezoidal shape, even though they are chemically distinct, that is found in a wide variety of virus families sharing little to no host, size or amino acid sequence similarity.^{25,26} Mathematically, it has been recently proved that the trapezoidal shape is the perfect building block to tile a closed icosahedral surface of any capsid size.²⁷ Exploiting this important role of the coat protein shape and the observation that the driving forces for capsid assembly arise from weak non-covalent inter-protein interactions of 3–4 kcal/mol,^{28,29} we developed coarse-grained models (Fig. 1) that represent each coat protein as a trapezoidal structure with inter-protein interactions described by generalized short-ranged, weak and anisotropic attractions. Our coat protein models represent each protein subunit as a set of 24 beads arranged in multiple layers confined in the trapezoidal geometry that accommodates the acute dihedral angle of 144° (between A-subunits or between B- and C-subunits along the five-fold symmetry axis) and a relatively flat angle of 170° (between B- and C-subunits along the six-fold symmetry axis). These models contain significantly more protein detail than models that represent each coat protein, capsomer or vertex as a single spherical bead with directional bondings or as a patchy particle suitable for the formation of T=1 capsids.^{14,16,17,19,20} Our geometric models are inspired by the approach of Rapaport *et al.*,^{12,30} who performed the first exploratory molecular dynamics simulations on capsid self-assembly of a polyhedral structure from trapezoidal subunits. Unlike these early simulations, which enforced many non-physical assembly rules ensuring that only full capsids would be formed and thus could not probe capture the spontaneous and reversible nature of capsid self-assembly, our simulations proceed without any specific built-in self-assembly rules. Coupled with discontinuous molecular dynamics, an extremely fast alternative to traditional molecular dynamics, our models allow us to simulate the spontaneous self-assembly of not only T=1 and T=3 icosahedral capsids but also non-icosahedral capsules, which requires significantly longer simulation times to observe and study.

We explore the assembly of T=1 and T=3 capsids, where T (triangulation number) denotes the number of subunits constituting an icosahedral asymmetric unit (T=1, 3, 4, 7, etc.) and the number of copies of the same coat protein (i.e., 60T proteins that are arranged as twelve pentamers and 10(T-1) hexamers at well-defined locations).² Protein subunits of only one conformation (i.e., pentamer-forming A-subunits) are needed to form T=1 capsids whereas subunits of three different conformations (i.e., pentamer-forming A-subunits and hexamer-forming B- and C-subunits) are needed to form T=3 capsids. Since all protein subunits (gene

products) are chemically identical,³¹ the switching of coat proteins between conformations that are poised to form either a pentameric or hexameric capsomer in $T \geq 3$ systems must occur during assembly. High-resolution *in vitro* studies,³² in which robust assembly by capsid subunits readily occurs to reconstitute empty icosahedral capsids with morphology indistinguishable from those virus particles assembled *in vivo*, reveal that the variations in protein conformations involve an alteration between order and disorder of the flexible regions located near the N- and C-termini, which are commonly referred to as “molecular switches”.^{33,34} When ordered, these arms interdigitate with their neighbors serving as wedges between proteins that are arranged in a hexamer; therefore, the angle between these wedged proteins along the six-fold symmetry axis is relatively flat whereas the angle between non-wedged proteins along the five-fold symmetry axis is relatively sharp. The exact nature of how this conformational switching mechanism (also known as autostery) takes place in deciding the appropriate conformation that is needed to be added onto the growing capsid at the right location is not well understood. Since the switching timescale is unknown but is local in nature, it may be anticipated to be facile compared to the timescales of protein assembly. Consequently, in constructing our models, we assume that a switching step is much faster than an assembly step and thus our simulations proceed from a pre-equilibrium of pentamer-templated and hexamer-templated protein units. Given the quasi three-fold symmetry of $T=3$ viral capsids, this gives rise to equal numbers of subunits in three predetermined conformations. Although the molecular switch is not explicitly represented in our models, its effect in flattening the dihedral angle between B- and C-subunits in hexamers is taken into account. Moreover, in order to account for the the role of autostery role in regulating the addition of a subunit of a certain conformation onto the growing structure during assembly at the right location, we impose specific interactions between certain subunits at the appropriate interfaces. A-subunits are allowed to form interfacial interactions at the five-fold symmetry axis whereas B- and C-subunits at the six-fold symmetry axis, and the three subunits are allowed to form interfacial interactions at the three-fold symmetry axis. In effect, imposing specific interactions allows the formation of pentamers and hexamers at the right locations; for instance, when a pentamer of A-subunits is formed first, only B- and C-subunits are added at the surrounding locations so that only hexamers can be grown.

We utilize molecular dynamics simulations to analyze molecular-level kinetic events that are difficult or impossible to explore experimentally during the self-assembly process. Molecular dynamics simulations were carried out at constant temperature in a fixed volume with periodic boundary conditions starting from a random configuration of exclusively monomers (i.e., A-subunits in $T=1$ systems and A-, B- and C subunits in $T=3$ systems) as the building blocks. Since the starting building blocks of many virus assembly systems have been shown to be dimeric³⁵ or trimeric,²⁴ we also performed simulations starting from a stoichiometric mixture of dimers that are derived from A-B subunits and C-C subunits, which are rigidly coupled together at the two-fold interface. Similarly, we also performed simulations starting from a configuration of trimers that are derived from A-, B- and C-subunits, which are rigidly coupled together at the three-fold interface. We examine and contrast the kinetics and thermodynamics governing spontaneous capsid self-assembly in $T=1$ and $T=3$ capsid systems under optimal conditions in which the icosahedral symmetry rule is obeyed. Moreover, we explore suboptimal conditions where the icosahedral symmetry rule breaks resulting in the formation of aberrant structures other than canonical icosahedral capsids. Based on these insights, we formulate strategies for the controlled assembly of icosahedral capsids and other non-icosahedral capsules by fine-tuning the conditions and redesigning the coat proteins.

Methods

Our coarse-grained models (Fig. 1), which capture the basic trapezoidal shape of a typical coat protein from $T=1$ and $T=3$ crystal structures, are used with discontinuous molecular dynamics

(DMD),³⁶ an extremely fast alternative to traditional molecular dynamics. In the DMD simulation algorithm all potentials are discontinuous, i.e., based on hard-sphere or square-well interactions. The solvent is modeled implicitly in the sense that inter-protein interactions represent solvent renormalized interactions. The excluded volume of each pseudo-atom is modeled using a hard-sphere potential. Covalent bonds are maintained between adjacent pseudo-atoms by imposing hard-sphere repulsions whenever the bond lengths fluctuate outside of the range $(1-\delta)L$ and $(1+\delta)L$, where L is the bond length and δ represents bond fluctuation and is 20% of L . These fluctuations mimic the conformational flexibility anticipated for individual protein subunits. Inter-actions between attractive sites are represented by a square well of depth ϵ and range 1.5σ , where σ is the pseudo-atom diameter. For details of the DMD methodology, see papers by Alder and Wainwright³⁶ and Smith, Hall, and Freeman.³⁷

The results presented in this paper are from averages over 100 independent simulations at each condition unless specified otherwise. Each $T=1$ simulation was performed on a system containing 1500 protein subunits, which can form up to 25 complete $T=1$ capsids. Each $T=3$ system contains either 1800 monomeric subunits, 900 dimeric subunits or 600 trimeric subunits, which can form up to 10 complete $T=3$ capsids. Each system was simulated over a wide range of temperatures and simulation box sizes (protein concentrations). The resulting protein concentrations were estimated to be $c=21.7$, 43.25, 86.5 and $173\mu\text{M}$ using the average dimension of coat proteins in crystal structures of small viral capsids.³⁸ Our range of temperatures between $T^*=0.30$ and 0.50 is equivalent to 264 and 440K, respectively, estimated using the value of -3.5 kcal/mol as the average protein inter-subunit interaction free energy as extracted from experiments on Hepatitis B virus capsids²⁸ (-3 to -4 kcal/mol) and cowpea chlorotic mottle virus capsids²⁹ (about -3 kcal/mol). Although our systems are relatively large, containing over 36,000 pseudo-atoms for the $T=1$ model or 43,200 pseudo-atoms for the $T=3$ models, they explore the complete assembly process starting from a random configuration and arriving at an equilibrium configuration that exhibits a variety of ordered and disordered structures, depending on the simulation conditions. The simulation time required to see the first complete capsid was relatively short, involving only a few days on a single-processor workstation. However, each system was simulated for a long period of time (about 30 days of cpu time) until the ensemble average of the total potential energy varied by no more than 2.5% during the last three-quarters of each simulation run.

Results

Observation of icosahedral capsids and non-icosahedral capsules at near-optimal conditions

Under the optimal conditions of protein concentration ($86.5\mu\text{M}$) and temperature (308K), assembly efficiency approaches 95% for the 60-subunit, $T=1$ system. Similar conditions produce a lower yield of $<80\%$ for 180-subunit, $T=3$ capsids; rotation function analysis³⁹ verifies that these complete capsids (Fig. 2A) are indeed icosahedral (SI, Fig. 7). It is remarkable that these precise structures spontaneously form given that assembly of complete capsids, even though an enthalpically driven process, is kinetically hindered. In previous studies, we demonstrated that although the early assembly steps take place on a downhill free-energy landscape, insertion of the final subunits is slow and can be rate-limiting.¹⁸ Moreover, one misstep, e.g., the formation of a single non-native protein-protein interface out of 150 ($T=1$) or 450 ($T=3$) native interfaces, can derail the assembly and cause the system to form other non-idealized structures.

Self-assembly of icosahedral capsids is strongly dependent upon the assembly conditions. In fact, slightly lowering the temperature from 308K to 290K at $86.5\mu\text{M}$ leads to polymorphism in $T=1$ assembly studies. At this temperature, in addition to icosahedral capsids, we observe a variety of non-icosahedral capsules that are enclosed and highly symmetric (Fig. 2B–G).

Similar to icosahedral capsids, non-icosahedral capsules contain 12 pentamers (required for structure closure) that are evenly distributed around each structure. The capsules are larger than the icosahedral T=1 capsids, containing a precise number of $N=60+6D$ monomeric subunits where D is the number of hexameric “dislocations”. The dislocations are comprised of kinetically trapped pentamer-templated A-subunits in a hexavalent environment. The relative position and number of hexameric dislocations define each capsule type: D-values of 2, 3, 4, 5 and 6 correspond to oblate, angular, twisted, tubular and prolate capsules respectively. The oblate and twisted capsules are semi-spherical whereas angular, tubular and prolate capsules are elongated. We also observe the formation of non-symmetric structures called conical capsules (Fig. 2I), where 12 pentamers are unevenly distributed in the structure (i.e., five pentamers are on one end of the structure while seven are localized on the other end).

The same sort of structural polymorphism is also observed in the T=3 system at 86.5 μM and 308K (Fig. 2). Besides icosahedral capsids, non-icosahedral yet completely enclosed structures such as oblate (Mov. S1), angular (Mov. S2), twisted (Mov. S3) and tubular (Mov. S4) capsules containing $N=180+18D$ monomeric subunits, where D-values are 2–5, are observed. The dislocations are comprised of kinetically-trapped pentamer-templated A-subunits in a strained conformation where at least one subunit makes sub-optimal interactions with its neighbors: in each hexameric dislocation A-subunits make fewer interactions than a spontaneously formed hexamer of B- and C-subunits. Pentameric dislocations containing B- and C-subunits are also observed during self-assembly of non-icosahedral capsules; however, they anneal to hexamers in the final non-icosahedral capsules. Although formed in significantly smaller numbers than icosahedral capsids, non-icosahedral capsules are consistently observed in our simulations in which the building blocks of the coat proteins are all either monomers, dimers or trimers (Fig. 2). Non-icosahedral capsules are relatively stable even upon heating up to 381K. However, since they are more susceptible to structural disruptions because of the strained conformations of the hexameric dislocation, they dissemble more readily than their icosahedral counterparts at higher temperatures (SI, Fig. 8). Similar to the T=1 system, we also observe the self-assembly of conical capsules that share the same morphology.

Caspar and Klug in their seminal work on the principle of quasi-equivalence² predicted that the most likely aberrant structures of lower symmetry, yet with similar diameter and surface structure to icosahedral virus particles, to be tubular capsules. The formation of icosahedral capsids and tubular capsules has also been predicted to occur by Bruinsma *et al.*⁹ and observed experimentally in various *in-vitro* assembly systems such as T=3 plant viruses (brome mosaic⁴⁰ and cowpea chlorotic mottle virus⁴¹), T=7 viruses (polyomavirus,⁴² simian virus 40⁴³ and papillomavirus⁴⁴) and bacteriophage $\phi 29$.⁴⁵ Our T=1 angular and tubular capsules are identical to oblong flock house virus structures, which were experimentally observed by Dong *et al.*²⁴ and confirmed by Chen *et al.*²³ Oblate capsules have been seen in T=1 alfalfa mosaic virus by Cusack *et al.*⁴⁶ Twisted capsules, to our knowledge, are reported here for the first time; however, we suspect that they are often mistakenly identified as either tubular capsules or enlarged icosahedral capsids because of their similar size and shape. Conical capsules and a variety of spherical and cylindrical structures have also been experimentally observed in retroviral capsids including HIV-1 by Ganser-Pornillos *et al.*⁴⁷ and examined theoretically with simple continuum elastic theory by Nguyen *et al.*²² Our findings taken with those just noted reinforce the notion that structural polymorphism is inherent to the nature of assembled coat proteins and that condition-dependent kinetic mechanisms play an important role in determining the various structures of viral capsids.

Kinetic mechanisms of icosahedral capsid and non-icosahedral capsule self-assembly

The mechanisms for the self-assembly of T=1 and T=3 icosahedral capsids at 86.5 μM and 308K are remarkably different from one another. The formation of T=1 capsids is governed

by elementary kinetics via addition of predominately monomers and occasionally small oligomers onto growing structures, as indicated in Fig. 3A by the continuity of the particle growth versus simulation time curves for individual simulations. Even though the coat protein concentration is relatively high compared to that utilized in experiments, the data in Fig. 3A indicates that there are three distinctive phases of assembly involving nucleation, rapid growth and completion. Accumulated data from 100 simulations under the same conditions shown in Fig. 4A indeed confirm the nucleation aspect of capsid assembly. Fig. 4A shows that the early stage of self-assembly of T=1 capsids is dominated by low order aggregates. Once nuclei are formed, such intermediates are transient and proceed rapidly to form aggregates of larger size and eventually complete capsids. At the end of each simulation, there is a partitioning between complete capsids and free monomers. This finding is similarly observed in light scattering and size exclusion chromatography experiments that monitor capsid assembly^{8,48,49} and the two generic thermodynamic-kinetic models that were developed by Zlotnick and coworkers.^{8,50,51} Moreover, the kinetic mechanism for the T=1 capsid self-assembly that is observed in our simulations is also seen in simplified model studies by Zhang and Schwartz,¹⁶ Hagan and Chandler¹⁷ and our previous studies¹⁸ in which two coarse-grained models of increasing detail, representing either three coplanar proteins as a capsomer or an individual protein as a structural unit. However, there is a significant difference between our results¹⁸ and those observed from the single-bead studies^{16,17} during the last stage of capsid assembly which involves the addition of the final subunits into the growing structure. Protein subunits that are modeled as single spherical beads are not sterically hindered and thus are more easily inserted into the growing structure whereas our protein subunits that are modeled as trapezoidal structures have to adjust themselves to a specific orientation relative to the capsid before the insertion step. Because of such high entropic costs, the amount of time needed to add each of the final subunits substantially increases as the growing structure approaches completion.

In contrast, T=3 icosahedral capsid self-assembly utilizes a mechanism in which preformed aggregates of intermediate sizes combine. The large jumps in the growth curves in Fig. 3B indicate that aggregates ranging from 10 to 40 proteins join the growing structures. Once these large aggregates combine, continued growth occurs at multiple points simultaneously with separate long, relatively flexible, “arms” that later merge to close the structure (Fig. 5A). This structural flexibility and collapse of conjoining faces can give rise to the formation of the hexameric dislocations that lead to non-icosahedral capsules. Kinetic data from 100 simulations show that self-assembly in the T=3 system involves a relatively large distribution of intermediates of varied sizes (indicated as small red dots in Fig. 4B), which collapse onto one another to form either icosahedral capsids or aberrant capsules. At the end of each simulation, there is no clear partitioning between complete capsids and other species since complete capsids, large capsules (predominately oblate capsules), intermediates of various aggregate sizes (at a low population) coexist with the free monomers.

Self-assembly of non-icosahedral capsules in both T=1 and T=3 systems involve the critical early step of forming hexameric dislocations, which arise during the collapse of intermediates of significant sizes containing non-complementary edges. For example, the formation of an oblate capsule (Fig. 5B) is initiated by the collapse of two isolated oligomers containing non-complementary edges, resulting in a trapped hexamer of A-subunits, which propagates the formation of a flat surface containing six pentamers of A-subunits in a ring rounding off at the edge. This process is replicated in reverse in forming the other symmetric half by first growing another ring containing six pentamers forcing B- and C-subunits at the growing edge to be on the same plane so that another hexamer of A-subunits is filled in to enclose the structure. In general, the critical step of forming the first hexameric dislocation instigates a series of coordinated events that yield additional hexameric dislocations, whose locations relative to the initial hexamers determine whether the growing structure can form an enclosed ordered structure or open misaggregate. Specifically, the formation of extra hexameric dislocations at

well-defined positions creates non-icosahedral capsules whereas the formation of randomly located dislocations leads to open misaggregates lacking any global symmetry. Additional hexameric dislocations are often formed to relieve the strain caused by the disjointed junctions between concave surfaces and flat initial hexameric dislocations by pulling an existing pentamer apart to insert another A-subunit or by filling in a flat hole to enclose a growing structure.

Interestingly, even though different simulations proceed from configurations of either all monomers, dimers or trimers of the coat proteins, the results from these simulations exhibit the same behavior not only in the endpoint structures and their population distributions but also in the kinetic mechanisms (data not shown). This implies that when the initial supply of coat proteins is limited, as long as subunits are started as either individual proteins or uniform capsomers of the same aggregate size, self-assembly of both icosahedral capsids and non-icosahedral capsules proceed via the same kinetic mechanisms, which involves both sequential addition of individual building blocks and condensation of preformed intermediates.

Towards controlled self-assembly by searching for the optimal assembly condition

Since the rapid establishment of a high population of partial structures appears to lead to the formation of large non-icosahedral capsules via the collapse-growth mechanism, alternative conditions for particle growth that minimizes the population of large partial structures and thus reduces the probability of forming structures with kinetically-trapped hexameric dislocations will likely increase the yield of T=3 icosahedral particles. Assembly initiated from a lower concentration of coat protein ($21.7\mu\text{M}$) and at lower temperature (290K) (Fig. 3C) appears to limit the rapid growth of multiple large aggregates and yield monomer addition driven growth.⁸ Indeed, by shifting to an alternative region in the phase diagram for particle growth, we find a higher population of icosahedral capsids assembled (SI, Fig. 10) albeit following significantly longer incubation periods. This points to the control of assembly conditions as a key strategy for the control of assembly products and the improvement of laboratory and manufacturing yields, which may serve as an instrumental means in increasing the potency of cervical cancer vaccines.⁶

Observation of misassembled aggregated or “monster particles” at extreme conditions

Increased protein concentration and lower temperatures, at the other extreme, e.g., $173\mu\text{M}$ and 264K, produce misassembled aggregates in simulations of both T=1 and T=3 systems. In this case, assembly is initiated too many times yielding an overabundance of nuclei that cannot proceed to completion as capsids and instead collapse to form non-native contacts thereby creating oversized misassembled aggregates.^{8,17,18,52} This specific situation in which kinetic pathways are pertinent to self assembly is an example of a generally appreciated concept.⁵³ In particular, a completed self-assembled cluster is composed of many units, and there are many pathways by which they can cluster. When intermediate structures initially form, they must have time to organize into an arrangement that is amenable to further assembly. If there is insufficient time to establish a local equilibrium, mistaken assemblies will occur when binding energies are too large or subunit concentrations are too high. In such cases, thermal agitation cannot anneal the structure before another subunit attaches to the cluster. This phenomenon was first analyzed by Zhang and Schwartz¹⁶ and Hagan and Chandler,¹⁷ and was further illustrated by Nguyen *et al.*¹⁸ More recently, a means of quantifying the phenomenon through measurements of fluctuation-dissipation ratios has been proposed.⁵⁴

Generally, we observe that the misassembled aggregates, also known as “monster particles”, emerging from the T=1 systems differ from those in T=3 systems (Fig. 6A–B). The T=1 monster particles tend to contain a few partial capsids and are relatively enclosed. T=3 misassembled aggregates have a more complex architecture that resembles the spiral structures

of bacteriophage P22 monster particles.⁵⁵ This diversity is a direct consequence of the inherent difference in the overall structure of the T=1 subunits, which form only pentameric “concave” capsomers, and T=3 subunits which assemble as a competition between these same pentameric “concave” capsomers and “flat” hexameric capsomers. This competition, and the fact that T=3 subunits can become kinetically trapped at low temperatures in helix-like clusters with more than six B- and C-subunits arranged alternately in spiral motifs, gives rise to additional non-spherical appendages hanging off the sides of the T=3-derived monster particles.

Coat protein engineering

The geometric differences in our structural units mirror the small variance observed in the structure of proteins in pentameric and hexameric capsomers from many icosahedral capsid crystal structures of various viruses (ranging from T=3 to 13).³⁸ In fact, for capsids without any apparent inter-subunit holes or overlaps, the structures of the same protein in either pentameric or hexameric positions often differ by only about 1.0Å RMSD in C_α atoms. These subtle conformational differences can be exploited in therapeutic development to interfere with the self-assembly of viable viral capsids. For example, when A-subunits are replaced by the averaged structure of A-, B- and C-subunits, meaning that the A-subunit’s ability to form pentamers is significantly reduced while its ability to form hexamers is enhanced, assemblies at 86.5μM and 308K are large, open, layered and disordered aggregates containing pentamers and hexamers at random locations (Fig. 6C). In this case, too many missteps occur for the growing structures to recover. Moreover, replacing the A-subunits by the averaged structure of the B- and C-subunits, to produce strongly biased hexameric capsomer forming species, results in assembly endpoints that are large, open and ordered aggregates containing only hexamers (Fig. 6D). These sheet-like aggregates grow and wrap around to form open-ended nanotubes, which could serve as a source of new biomaterials. These results provide the potential for direct testing and may already be confirmed in recent experiments by Stray *et al.*⁵² who show that one can alter the yield of capsid particles by binding a “wedge” between the protein interfaces and hence artificially force specific interfaces into the assembly process.

Discussion

In systems comprised of both pentameric and hexameric capsomers, i.e. for T≥3, the switching of coat proteins between conformations that are poised to form either a pentameric or hexameric capsomer must occur during assembly, since all protein subunits are chemically identical.³¹ We account for this switching mechanism in our T=3 assembly studies by (1) explicitly representing protein subunits in a pre-equilibrium of the three predetermined conformations and (2) imposing specific interactions between protein subunits at appropriate interfaces. This approach is able to regulate the formation of pentamers and hexamers at well-defined locations to yield icosahedral capsids based only on the geometric shapes and protein-protein interactions of the subunits. However, when the pre-equilibrium is shifted to favor pentamer-templated protein A-subunits, the proclivity for five-to-six (and higher) symmetric dislocations increases resulting in the formation of oblate and other aberrant structures (SI, Fig. 10). One notable type of structure observed under these conditions are monster particles that resemble oblate and other non-icosahedral capsules; however, they are not completely enclosed due to the presence of too many A-subunits that grow on each side of the opening.

When accounting for the conformational switching mechanism, we assume that it must occur in the bulk as a monomeric species, possibly instantly before the protein subunit is added onto the growing structure. Once assembled, even if autostery could take place between protein subunits that are in the aggregated state, spontaneous switching between a pentamer and a hexamer is unlikely since such highly cooperative process that involves five to six protein subunits would require overcoming a large energy barrier for either removing one extra protein

subunit when a hexamer is converted into a pentamer or adding an additional protein subunit when a pentamer is converted into a hexamer. In the first case, not only does one subunit have to be removed from the hexamer but the protein subunits that are attached to the excised subunit must be removed as well (Fig. 5B); therefore, the probability that such a transformation would occur decreases as the number of the attached subunits increases. Consequently, once sizable intermediate aggregates collapse into one another, the combined structure is significantly stable as it keeps growing into non-icosahedral capsules. In the latter case, adding an additional protein subunit onto a closed pen-tamer requires pulling one of its interfaces apart by breaking not only those interactions present in the pentamer but also those interactions of the protein subunits that are attached to the pentamer along that interface as well. Therefore, spontaneous switching between a pentamer and a hexamer in the aggregated state is unlikely to occur unless with the help of auxiliary machinery or drastic changes in the environmental conditions.

The self-assembly of non-icosahedral capsules is a consequence of an off-pathway mechanism that is prevalent under non-optimal conditions rather than the result of improperly accounting for autostery. In $T=1$ systems, where the conformational switch mechanism is unnecessary in capsid self-assembly since $T=1$ capsids contain only one conformation of the coat proteins, yet the same type of non-icosahedral capsules as those observed in $T=3$ systems emerges. Similarly, in-vitro experiments by Dong *et al.*²⁴ reveal that when the molecular switch of the $T=3$ flock house virus is deleted, icosahedral capsids of $T=1$ size are observed in addition to multiple types of oblong capsules that are larger than $T=1$ capsids and smaller than $T=3$ capsids. Analysis of these oblong capsules indicates that they share the same morphology and size as those observed from our simulations. These results suggest that the conformational switch mechanism would not be able to prevent the self-assembly of polymorphic structures in $T=3$ systems once hexameric dislocations have been kinetically trapped.

Conclusion

The combination of coarse-grained models, which capture the essential geometry and interaction details of real coat proteins yet are simple enough to allow the dynamical simulation of systems containing many such proteins, with discontinuous molecular dynamics,³⁶ have been used to simulate the spontaneous formation of supramolecular structures of large $T=1$ and $T=3$ viral capsids (without the genomic contents) initiated from random configurations of only protein monomers, dimers or trimers. Our findings demonstrate that structural polymorphism in the endpoint structures is an inherent property of coat proteins, is independent of the morphology of constituent subunits, and arises from condition-dependent kinetic mechanisms that are determined by initial assembly conditions related to protein concentration and temperature. Based on the ubiquitous nature of non-icosahedral capsules observed in our simulations, we predict that such capsules also occur in $T>3$ systems having precise numbers of $N=60 \cdot T + 6 \cdot T \cdot D$ subunits. We also show that the assembly of icosahedral capsids and other enclosed non-icosahedral capsules can be controlled by fine-tuning assembly conditions. Moreover, we illustrate that conformational switching within individual coat proteins controls their ability to form pentameric capsomers, altering the formation of enclosed structures and interfering with the self-assembly of viable virus particles. Simulation studies as described here and continuing efforts to examine capsid self-assembly in specific virus systems provide new tools to inform potential strategies in antiviral development, protein design and the engineering of novel biomaterials.

The genomic content and/or other auxiliary components such as scaffold proteins play in important role in the self-assembly of full viruses. They may serve as a template or a means to control the formation of icosahedral capsids and thus prevent the formation of polymorphic capsules. The models that we have developed in this study can be used to further examine the

role of the genomic content and scaffold proteins in improving the fidelity of structural morphology in viral capsids.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank Dennis C. Rapaport, Jack E. Johnson and Robijn Bruinsma for helpful discussions. This work was supported by the National Institutes of Health under grant number RR12255 and the National Science Foundation under grant number PHY0216576.

References

1. Crick FH, Watson JD. *Nature* 1956;177:473–475. [PubMed: 13309339]
2. Caspar DLD, Klug A. *Cold Spring Harb Symp Quant Biol* 1962;27:1–24. [PubMed: 14019094]
3. Douglas T, Young M. *Science* 2006;312:873–875. [PubMed: 16690856]
4. Koutsky LA, Ault KA, Wheeler CM, Brown DR, Barr E, Alvarez FB, Chiacchierini LM, Jansen KU. *N Engl J Med* 2002;347:1645–1651. [PubMed: 12444178]
5. Shank-Retzlaff M, et al. *Hum Vaccin* 2005;1:191–7. [PubMed: 17012876]
6. Shi L, Sings HL, Bryan JT, Wang B, Wang Y, Mach H, Kosinski M, Washabaugh MW, Sitrin R, Barr E. *Clin Pharmacol Ther* 2007;81:259–64. [PubMed: 17259949]
7. Wales D. *Chem Phys Lett* 1987;141:478–484.
8. Zlotnick A, Johnson JM, Wingfield PW, Stahl SJ, Endres D. *Biochemistry* 1999;38:14644–52. [PubMed: 10545189]
9. Bruinsma RF, Gelbart WM, Reguera D, Rudnick J, Zandi R. *Phys Rev Lett* 2003;90:248101. [PubMed: 12857229]
10. Twarock R. *J Theor Biol* 2004;226:477–482. [PubMed: 14759653]
11. Zandi R, Reguera D, Bruinsma RF, Gelbart WM, Rudnick J. *Proc Natl Acad Sci U S A* 2004;101:15556–60. [PubMed: 15486087]
12. Rapaport DC. *Phys Rev E Stat Nonlin Soft Matter Phys* 2004;70:051905–051917. [PubMed: 15600654]
13. Freddolino PL, Arkhipov AS, Larson SB, McPherson A, Schulten K. *Structure* 2006;14:437–449. [PubMed: 16531228]
14. Van Workum K, Douglas JF. *Phys Rev E Stat Nonlin Soft Matter Phys* 2006;73:031502. [PubMed: 16605527]
15. Hemberg M, Yaliraki SN, Barahona M. *Biophys J* 2006;90:3029–42. [PubMed: 16473916]
16. Zhang T, Schwartz R. *Biophys J* 2006;90:57–64. [PubMed: 16214864]
17. Hagan MF, Chandler D. *Biophys J* 2006;91:42–54. [PubMed: 16565055]
18. Nguyen HD, Reddy VS, Brooks CL III. *Nano Lett* 2007;7:338–344. [PubMed: 17297998]
19. Wilber AW, Doye JP, Louis AA, Noya EG, Miller MA, Wong P. *J Chem Phys* 2007;127:085106. [PubMed: 17764305]
20. Chen T, Zhang Z, Glotzer SC. *Proc Natl Acad Sci U S A* 2007;104:717–22. [PubMed: 17215354]
21. Twarock R. *Bull Math Biol* 2005;67:973–87. [PubMed: 15998491]
22. Nguyen TT, Bruinsma RF, Gelbart WM. *Phys Rev Lett* 2006;96:078102. [PubMed: 16606144]
23. Chen T, Glotzer SC. *Phys Rev E Stat Nonlin Soft Matter Phys* 2007;75:051504. [PubMed: 17677070]
24. Dong XF, Natarajan P, Tihova M, Johnson JE, Schneemann A. *J Virol* 1998;72:6024–6033. [PubMed: 9621065]
25. Rossmann MG, Johnson JE. *Annu Rev Biochem* 1989;58:533–73. [PubMed: 2673017]
26. Harrison SC. *Seminars in Virology* 1990;1:387–403.
27. Mannige R, Brooks CL III. *Phys Rev E* 2008;77:051902.

28. Ceres P, Zlotnick A. *Biochemistry* 2002;41:11525–11531. [PubMed: 12269796]
29. Johnson JM, Tang J, Nyame Y, Willits D, Young MJ, Zlotnick A. *Nano Lett* 2005;4:765–770. [PubMed: 15826125]
30. Rapaport DC, Johnson JE, Skolnick J. *Comp Phys Comm* 1999;121:231–235.
31. Rossmann MG. *Virology* 1984;134:1–11. [PubMed: 6710869]
32. Harrison SC. *Trends in Biochemical Sciences* 1984;9:345–351.
33. Speir JA, Munshi S, Wang G, Baker TS, Johnson JE. *Structure* 1995;3:63–78. [PubMed: 7743132]
34. Zhao X, Fox JM, Olson NH, Baker TS, Young MJ. *Virology* 1995;207:486–94. [PubMed: 7886952]
35. Sorger PK, Stockley PG, Harrison SC. *J Mol Biol* 1986;191:639–658. [PubMed: 3806677]
36. Alder BJ, Wainwright T. *J Chem Phys* 1959;31:459–466.
37. Smith SW, Hall CK, Freeman BD. *J Com Phys* 1997;134:16–30.
38. Shepherd CM, Borelli IA, Lander G, Natarajan P, Siddavanahalli V, Bajaj C, Johnson JE, Brooks CL III, Reddy VS. *Nucleic Acids Res* 2006;34:D386–D389. [PubMed: 16381893]
39. Tong LA, Rossmann MG. *Acta Crystallogr A* 1990;46:783–92. [PubMed: 2174243]
40. Bancroft JB, Bracker CE, Wagner GW. *Virology* 1969;38:324–35. [PubMed: 5784855]
41. Adolph KW, Butler PJ. *J Mol Biol* 1974;88:327–41. [PubMed: 4452998]
42. Salunke DM, Caspar DL, Garcea RL. *Biophys J* 1989;56:887–900. [PubMed: 2557933]
43. Kanesashi SN, Ishizu K, Kawano MA, Han SI, Tomita S, Watanabe H, Kataoka K, Handa H. *J Gen Virol* 2003;84:1899–905. [PubMed: 12810885]
44. Zhao Q, Guo HH, Wang Y, Washabaugh MW, Sitrin RD. *J Virol Methods* 2005;127:133–40. [PubMed: 15894387]
45. Fu CY, Morais MC, Battisti AJ, Rossmann MG, Prevelige JPE. *J Mol Biol* 2007;366:1161–73. [PubMed: 17198713]
46. Cusack S, Oostergetel GT, Krijgsman PC, Mellema JE. *J Mol Biol* 1983;171:139–55. [PubMed: 6655690]
47. Ganser-Pornillos BK, von Schwedler UK, Stray KM, Aiken C, Sundquist WI. *J Virol* 2004;78:2545–52. [PubMed: 14963157]
48. Zlotnick A, Aldrich R, Johnson JM, Ceres P, Young MJ. *Virology* 2000;277:450–456. [PubMed: 11080492]
49. Casini GL, Graham D, Heine D, Garcea RL, Wu DL. *Virology* 2004;325:320–327. [PubMed: 15246271]
50. Zlotnick A. *J Mol Biol* 1994;241:59–67. [PubMed: 8051707]
51. Endres D, Zlotnick A. *Biophys J* 2002;83:1217–1230. [PubMed: 12124301]
52. Stray SJ, Zlotnick A. *J Mol Recognit* 2006;19:542–8. [PubMed: 17006877]
53. Whitesides GM, Boncheva M. *Proc Natl Acad Sci U S A* 2002;99:4769–74. [PubMed: 11959929]
54. Jack RL, Hagan MF, Chandler D. *Phys Rev E* 2007;76:021119.
55. Earnshaw W, King J. *J Mol Biol* 1978;126:721–747. [PubMed: 370407]

Crystal structures of T=1 & T=3 capsids

T=1 & T=3 building blocks as monomers

T=3 building blocks as dimers or trimers

Figure 1.

Coarse-grained models of coat proteins as monomers for T=1 and T=3 systems; models of coat proteins as dimers and trimers are also employed for T=3 systems. Our monomer models capture the basic trapezoidal shape of a typical coat protein from crystal structures.³⁸ A T=1 icosahedral capsid is comprised of 60 coat proteins of the same type (A-subunit) arranged in 12 pentamers while a T=3 icosahedral capsid contains 180 coat proteins of three different types (A-, B- and C-subunits, in equal number) arranged as 12 pentamers of A-subunits and 20 hexamers of B- and C-subunits. Subunits of the A-type are templated for pentamer formation and B- and C-subunits are templated for hexamer formation. The dimer models represent capsomers of either A- and B-subunits or C- and C-subunits; the trimer models represent

capsomers of A-, Band C-subunits. The top and bottom layers of each subunit contain hard spheres providing volume exclusion. To account for autostery,³¹ specific inter-subunit interactions are allowed between pseudo-atoms of the same, non-white, color on similar middle layers. When two subunits are perfectly aligned at the correct angle forming a maximum number of four attractive interactions, the resulting interface is considered “native”; a “non-native” interface can occur when two subunits are locked at a wrong angle forming a fewer number of attractive interactions. Molecular dynamics simulations were carried out at constant temperature in a fixed volume with periodic boundary conditions without any built-in self-assembly rules starting from a random configuration of exclusively monomers, dimers or trimers as the building blocks.

Figure 2.

The population distribution obtained from 100 simulations for the T=1 system at $86.5\mu\text{M}$ and 290K and the T=3 system at $86.5\mu\text{M}$ and 308K: (A) complete icosahedral capsid, (B) oblate capsule, (C) angular capsule, (D) twisted capsule, (E) tubular capsule, (F) prolate capsule, (G) conical capsule, (H) partial capsid and (I) open mis-aggregate. The population distribution for supramolecular structures from T=3 systems containing coat proteins represented as either monomers (black), dimers (purple) or trimers (orange) is shown.

Figure 3.

The time-dependent aggregate size of species that grow into complete capsids. Three representative growth curves (red, blue, black) are illustrated for (A) a $T=1$ system at $86.5\mu\text{M}$ and 308K , (B) a $T=3$ system at $86.5\mu\text{M}$ and 308K and (C) a $T=3$ system at $21.7\mu\text{M}$ and 290K to show that condition-dependent growth yields kinetic mechanisms comprising predominately monomer addition (A and C) and large partial capsid collapse (B)

Figure 4.

The difference in the kinetic assembly mechanism for T=1 and T=3 systems under analogous conditions is shown by the probability of forming at least one aggregate of a certain size as a function of time. The results are obtained from 100 independent simulations at 86.5 μ M and 308K for (A) T=1 and (B) T=3 systems.

Figure 5.

Representative stages of aggregate growth of a (A) T=3 icosahedral capsid at and (B) oblate capsule at $86.5\mu\text{M}$ and 308K. The red arrow shows the first hexameric dislocation resulting from a collapse event during the self-assembly of the oblate capsule. If the hexameric dislocation was allowed to switch to a pentamer so the combined structure could grow into an icosahedral capsid, three extra protein subunits next the the red arrow have to be removed. Such transformation requires overcoming a relatively high free energy barrier and thus is an event of low probability.

Figure 6.

Snapshots of assemblies that are: (A) relatively enclosed monster particle obtained at $173\mu\text{M}$ and 264K from the $T=1$ system; (B) large and spiral monster particle obtained at $173\mu\text{M}$ and 264K from the $T=3$ system; (C) large, open, layered and disordered when A-subunits are replaced by the averaged structure of A-, B- and C-subunits; and (D) large, open and ordered when A-subunits are replaced by the averaged structure of B- and C-subunits at $86.5\mu\text{M}$ and 308K for the $T=3$ system.