

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/24278967>

Generalized Structural Polymorphism in Self-Assembled Viral Particles

ARTICLE *in* NANO LETTERS · JANUARY 2009

Impact Factor: 13.59 · DOI: 10.1021/nl802828v · Source: PubMed

CITATIONS

37

READS

22

2 AUTHORS:



[Hung D. Nguyen](#)
University of California, Irvine

29 PUBLICATIONS 750 CITATIONS

[SEE PROFILE](#)



[Charles L Brooks](#)
University of Michigan

190 PUBLICATIONS 13,649 CITATIONS

[SEE PROFILE](#)



NIH Public Access

Author Manuscript

Nano Lett. Author manuscript; available in PMC 2009 December 1.

Published in final edited form as:
Nano Lett. 2008 December ; 8(12): 4574–4581.

Generalized Structural Polymorphism in Self-Assembled Viral Particles

Hung D. Nguyen and Charles L. Brooks III

Department of Chemistry and Biophysics Program, 930 N. University Ave, University of Michigan, Ann Arbor, MI 48109, USA

Abstract

The protein shells, called capsids, of nearly all spherical viruses adopt icosahedral symmetry; however, self-assembly of such empty structures often occurs with multiple mis-assembly steps resulting in the formation of aberrant structures. Using simple models that represent the coat proteins pre-assembled in the two different predetermined species that are common motifs of viral capsids (i.e., pentameric and hexameric capsomers), we perform molecular dynamics simulations of the spontaneous self-assembly of viral capsids of different sizes containing 1, 3, 4 to 19 proteins in their icosahedral repeating unit ($T=1, 3, 4$ to 19 , respectively). We observe, in addition to icosahedral capsids, a variety of non-icosahedral yet highly ordered and enclosed capsules. Such structural polymorphism is demonstrated to be an inherent property of the coat proteins, independent of the capsid complexity and the elementary kinetic mechanisms. Moreover, there exist two distinctive classes of polymorphic structures: aberrant capsules that are larger than their respective icosahedral capsids, in $T=1-7$ systems; and capsules that are smaller than their respective icosahedral capsids when $T=7-19$. Different kinetic mechanisms responsible for self-assembly of those classes of aberrant structures are deciphered, providing insights into the control of the self-assembly of icosahedral capsids.

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, exhibits remarkable physical homogeneity *in vivo*. To function as infectious agents, virion particles are self-assembled reproducibly, reliably and efficiently, arriving at a monodisperse population of a particular capsid size that is encoded in the genome. Each capsid size comprises $60T$ copies of the coat protein (CP) arranged as twelve pentamers and $10(T-1)$ hexamers, where T (triangulation number) denotes the number of subunits constituting an icosahedral asymmetric unit ($T=1, 3, 4, 7, 9, 12, 13, 16, 19$, etc.) (Fig. 1A). During assembly, CP subunits are known to adopt several different conformations accommodating either fivefold or sixfold arrangements at well-defined positions.^{3,4} Factors regulating the selection of optimal CP subunit conformations to maintain physical homogeneity among assembled virions remain ambiguous. Nevertheless, high-resolution *in vitro* studies, in which robust assembly of CP subunits, in the absence of the genomic content and other supporting materials, readily occurs to reconstitute virus-like particles with morphology indistinguishable from those assembled *in vivo*.⁵ This assembly can produce not only icosahedral capsids, but also a range of non-icosahedral yet enclosed misassemblies such as elongated capsules.^{6,7} Since many beneficial applications of viral capsids are emerging in materials science and medicine,⁸ elucidating the

kinetic mechanisms and thermodynamic control of capsid self-assembly is a crucial prerequisite to advancing this area.

The spontaneous self-assembly of icosahedral capsids has been examined via many simulation studies,⁹⁻¹³ which provide invaluable insights into the kinetic mechanism of the process. However, except for a few theoretical studies,¹⁴⁻¹⁷ the spontaneous self-assembly of non-icosahedral capsules has not been broadly explained. We attempted to pursue this direction in an earlier study¹⁸ by first examining the optimal assembly of icosahedral capsids, which is strongly dependent upon the assembly conditions. We then searched for alternative structures of lower symmetry yet with similar stability that spontaneously formed under slightly non-optimal conditions. We used simple models, which realistically capture the geometric shape of each protein as a trapezoidal structure,¹⁹ for the self-assembly of T=1 and T=3 capsids. Our molecular dynamics simulations, which were performed with either monomeric, dimeric or trimeric CP subunits as the starting building blocks, revealed that viral CPs in T=1 and T=3 systems spontaneously assemble to, besides icosahedral capsids, an identical set of closed, non-spherical particles with $60T + 6DT$ proteins. D , which ranges between 2 to 6, is the number of hexameric dislocations, which occur as kinetically trapped hexameric arrangements of proteins in “pentameric positions” on the icosahedral lattice. The composition of this underlying phase space appears to be independent of both T and the elemental mechanism of capsid assembly. Hexameric dislocations, which are caused by the collapse events involving sizable intermediates, tend to take place at slightly higher protein concentrations and lower temperatures than those that yield predominantly icosahedral capsids. Furthermore, using the assembly conditions as a means for self-assembly control, a simple phase diagram was mapped out indicating the optimal conditions at which the kinetic mechanism responsible for efficient self-assembly of icosahedral capsids was deciphered to involve the addition of monomeric species.

In this study, we examine structural polymorphism further in larger capsid systems (e.g., up to T=19) by extending our models (Fig. 1B) to represent each pentameric CP capsomer as a pentagonal structure and each hexameric CP capsomer as a hexagonal structure. We determine structural polymorphism as a function of the T capsid system by first ascertaining the presence of the same aberrant particles as observed in the T=1 and T=3 systems and then examine the structural morphology of non-icosahedral capsules in different capsid systems. Moreover, we elucidate the kinetic mechanisms that are responsible for the formation of aberrant particles as a function of the assembly conditions (protein concentration and temperature) in order to formulate strategies for the controlled assembly of structures of any architecture.

Methods

Our pentameric/hexameric models (Fig. 1B) mimic the kinetic pathways and assembly species of a few known virus systems such as HK97 capsids, which have been shown to assemble from pentamers and hexamers.²⁰ Bruinsma and coworkers have demonstrated that two capsomers that are morphologically different (e.g. pentameric and hexameric) are necessary to produce structures of exclusively icosahedral symmetry when assembled on a spherical surface.^{14,21} Otherwise, a broader class of structures of other symmetries are observed.^{14,21} This has also been demonstrated by Glotzer and coworkers.²² They found that, depending upon the convexity constraint, identical units assemble into clusters of different sizes, few of which correspond to a range of icosahedral capsids. Our models were designed geometrically to ensure that every two capsomers are aligned at an appropriate curvature so they were able to form the maximum number of inter-capsomer interactions and the resulting capsid is the structure of minimum free energy. To ensure that these protein models are not biased towards the formation of capsids of any particular size, the main design requirements are (1) all five hexamers surrounding a pentamer form a structure with a tight concave curvature and (2) all six hexamers surrounding

a hexamer form a tight flat surface. The large spheres are repulsive providing volume exclusion. The small red spheres are attractive providing anisotropic inter-capsomer interactions if they are located on similar layers and belong to the interfaces that are specified according to the T capsid system as suggested by Kerner.²³ For example, the T=1 system, without hexamers, requires pentamers to be attractive to one another at any interface. Since pentamers are located further apart as T increases, each hexamer in T=3, 4, and 7 capsids is attractive to any pentamer at three, two, and one interfaces, respectively. Because some hexamers are completely distant from any pentamers in T \geq 9 capsids, the T=9 and 13 systems require one additional distinctive hexameric species whereas T=12, 16 and 19 systems require two additional distinctive hexameric species. In capsid systems with enantiomorphism (T=7, 13 and 19), both versions are considered by imposing different pentamer/hexamer interaction schemes.²³

These models are used with discontinuous molecular dynamics (DMD),²⁴ 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 between adjacent pseudo-atoms comprising the pentagonal capsomer and hexagonal capsomer are maintained by imposing hard-sphere repulsions whenever the bond lengths fluctuate outside of the range $l(1-\delta)$ and $l(1+\delta)$, where δ is the bond length representing the bond fluctuation and is 20% of l . These fluctuations mimic the conformational flexibility anticipated for individual protein subunits. Interactions between attractive sites are represented by a square well of depth ϵ and range 1.5σ , where σ is the pseudo-atom diameter. For each T system at a given assembly condition, 100 molecular dynamics simulations were carried out at constant temperature in a fixed volume with periodic boundary conditions. No built-in self-assembly rules were employed. Each simulation was performed on a system containing a mixture of 2000 pentameric and hexameric capsomers whose compositions are appropriately stoichiometrical based on the T capsid system. Each system was simulated at protein concentrations of 43.5 and 87 μM , which were estimated from the average dimension of CPs in crystal structures of small viral capsids,²⁵ and temperatures of 290 and 308K, which were estimated using a value of -3.5 kcal/mol as the average protein inter-subunit interaction free energy as extracted from experiments on Hepatitis B virus capsids²⁶ and cowpea chlorotic mottle virus capsids.²⁷ Although our systems are relatively large, containing over 70,000 pseudo-atoms, they explore the complete assembly process starting from a random configuration of capsomers 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 15 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 and Discussion

Observation of large non-icosahedral capsules in T=1-4 systems

In the systems of T=1, T=3 and T=4, efficiency is relatively high for the assembly of icosahedral capsids, whose symmetry is confirmed by rotation function analysis,²⁸ under optimal conditions (e.g., at a protein concentration of 43.5 μM and temperature of 308K); in fact, the icosahedral capsid yield approaches 90%. However, as the assembly conditions shift to an extreme condition of high protein concentration (173 μM) and low temperature (264K), misassembled aggregates, which are also known as “monster particles”, are predominately produced containing many non-native contacts.¹⁸ The morphology of these monster particles

is diverse. On the one hand, monster particles that are emerged from small capsid systems ($T \leq 3$) contain a few partial capsids and are relatively enclosed, resembling Turnip crinkle virus monster particles.²⁹ On the other hand, the particles that are observed from larger capsid systems ($T \leq 3$) have a more complex architecture, resembling the spiral structures of bacteriophage P22 monster particles.³⁰

When the assembly conditions shift from the optimal conditions to a higher protein concentration ($87\mu M$) or lower temperature (290K), structural polymorphism emerges in the form of icosahedral capsids and a variety of non-icosahedral capsules that are enclosed and highly symmetric (Fig. 2). In these simulations, the icosahedral capsid yield drops to 70% while the rest of the distribution belongs to non-icosahedral capsules whose sizes are inversely proportional to their yields. Similar to icosahedral capsids, non-icosahedral capsules contain 12 pentamers, which are required for structure closure and are evenly distributed around each structure. By contrast, these capsules are larger than icosahedral capsids and contain a precise number of $C = 12 + 10(T - 1) + DT$ capsomers, which is equivalent to the formula for the number of CP subunits observed in non-icosahedral capsules in $T=1$ and $T=3$ systems.¹⁸ Here D is the number of hexameric “dislocations”, which are comprised of six kinetically trapped hexamers surrounding a pentamer; therefore, their conformation is strained with the pentamer making sub-optimal interactions with at least one of its hexameric neighbors. The relative position and number of hexameric dislocations (i.e., wrong local arrangements on the icosahedral lattice) define the morphology of each capsule: D -values of 2, 3, 4, 5 and 6 correspond to oblate, angular, twisted, tubular and oblong capsules, respectively. The oblate and twisted capsules are semi-spherical whereas angular, tubular and oblong capsules are elongated. Unlike the icosahedral capsid, which is somewhat spherical and can be imagined in two dimensions (Fig. 1A) as having two rows containing ten pentamers and two pentamers on the top and bottom of the structure, each oblate capsule is relatively flat on the two ends where hexameric dislocations are positioned and the 12 pentamers are evenly distributed in two rows in the middle. This can be envisioned by adding another panel of four faces that are marked by the green arrow on the two-dimensional hexagonal template in Fig. 1A to the side and folding the surface up into a three-dimensional structure. At each end, six hexamers are trapped in a ring forming a central flat area as a hexameric dislocation. Instead of one hexameric dislocation at each end of the oblate capsule, each twisted capsule contains two hexameric dislocations at each somewhat flat end. Angular ($D=3$), tubular ($D=5$) and oblong ($D=6$) capsules have been well-characterized by in references.^{17,31} We also observe the formation of non-symmetric structures called conical capsules (Fig. 2), where the 12 pentamers are unevenly distributed in the structure (i.e., five pentamers are located on one end of the capsule with seven on the other end).

Tubular capsules have been observed experimentally in various in-vitro assembly systems such as $T=3$ plant viruses (brome mosaic⁶ and cowpea chlorotic mottle virus³²) and bacteriophage $\phi 29$.³³ Oblate capsules have been seen in $T=1$ alfalfa mosaic virus.³⁴ The transition between icosahedral capsids and tubular capsules has also been experimentally observed in cowpea chlorotic mottle virus³² and theoretically predicted to occur by Bruinsma *et al.*¹⁴ 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. Both of the spontaneously assembled prolate capsules that we observe ($D=3$ and $D=6$) appear to be identical to aberrant flock house virus structures which were experimentally observed by Dong *et al.*⁷ These structures were also confirmed in simulation studies by Chen and Glotzer¹⁷ who performed Monte Carlo simulations of equal-sized hard spheres representing either pentameric or hexameric capsomers; these spheres interact with an attractive square well on a prolate spheroid surface. Conical capsules and a variety of non-icosahedral yet spherical and cylindrical structures have also been

experimentally observed in retroviral capsids including HIV-1 by Ganser-Pornillos *et al.*³⁵ and examined theoretically with simple continuum elastic theory by Nguyen *et al.*¹⁵

Self-assembly of non-icosahedral capsules involves the critical early step of capsomers becoming kinetically trapped in hexameric dislocations, which arise during the collapse of sizable intermediates containing non-complementary edges. This step 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 disordered mis-aggregates. Therefore, non-icosahedral capsule self-assembly is a consequence of an off-pathway mechanism that is prevalent under non-optimal conditions. Once trapped, a hexameric dislocation remains stable since “switching” to the pentamer would require overcoming a large energy barrier in removing extra capsomers that are not only immediately involved in the dislocation but also those in the surrounding area. Furthermore, the trapped pentamer is loosely fitted in a hexavalent environment, thus it could add another coat protein subunit if monomeric proteins were freely available and thus become a hexamer, albeit still at a dislocation because its surrounding hexavalent environment is at a wrong position on the icosahedral lattice. This results in a more stable assembly as each protein subunit makes the most interactions with its neighbors. Consequently, non-icosahedral capsules are energetically equivalent with icosahedral capsids.

The formation of hexameric dislocations and thus self-assembly of non-icosahedral capsules as has emerged from our simulations can not be prevented by autostery, also known as the conformational switch mechanism, which is believed to regulate the exchange of capsid protein conformations for the formation of either pentameric or hexameric capsomers at well-defined locations on an icosahedral capsid.³⁶ High-resolution x-ray crystallographic studies 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”.^{3,37} When ordered, these arms interdigitate with their neighbors serving as wedges between proteins 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. By representing the CP subunits as either pentameric or hexameric capsomers, our simulations effectively proceed after the switching steps have occurred; therefore, autostery is no longer involved in our simulations. Moreover, once CP subunits are already in partially assembled states, the resulting pentameric and hexameric capsomers are unlikely to exchange spontaneously. As discussed earlier, such “switching” events are unlikely to occur since it requires overcoming a large energy barrier in removing or adding extra CP subunits once the capsomers that need to be “switched” are already situated in the middle of a large aggregate that is the result of a collapse event involving sizable intermediates.

In simulations of the T=1 capsid system, we also observe the assembly of geminate structures (Fig. 2) that resemble the characteristic twinned particles in Maize streak virus,³⁸ comprising two joined, incomplete T=1 icosahedral capsids and a total of 110 CP subunits arranged in 22 pentameric capsomers. Detailed structural analysis performed by Zhang *et al.* on Maize streak virus particles³⁹ suggests that the two incomplete T=1 capsids of a geminate particle merge at their openings, each missing a cluster of five proteins at a five-fold vertex. Thus we conjecture that the building blocks of geminate particle self-assembly are pentameric capsomers and the most likely scenario to produce a homogeneous population of such incomplete capsids should involve pentameric capsomers as the building blocks; otherwise, incomplete capsids with heterogeneous openings would be produced, which is not conducive to geminate particle self-assembly. Such was the case in our earlier study in which CP subunits are either monomeric, dimeric or trimeric; geminate particles were not able to form spontaneously. In addition, our other previous studies¹³ demonstrated that insertion of the final subunits is slow and can be

rate-limiting, although the early assembly steps take place on a downhill free-energy landscape, and the residence time for incomplete capsids with one building block missing is much longer than preceding growing steps, we reason that the availability of such incomplete capsids should facilitate the merging process.

Observation of both small and large non-icosahedral capsules in T=7 system

In addition to the same set of polymorphic capsules that are larger than icosahedral capsids as observed in T=1, T=3 and T=4 systems, the T=7 system provides another set of distinctive polymorphic capsules that are smaller than the native icosahedral capsids (Fig. 3) in both enantiomeric versions. Under optimal conditions, icosahedral capsids are grown by predominately adding capsomers sequentially. At protein concentrations that are slightly higher than the optimal conditions, many nuclei of aggregates appear and grow separately into sizable intermediates, which then collide and create structures with kinetically trapped hexameric dislocations that result in the formation of large non-icosahedral capsules such as oblate structures (Fig. 3B). At protein concentrations that are slightly lower than the optimal conditions, aggregates grow slowly into partial capsids, each of which tends to fold onto itself resulting in the formation of small non-icosahedral capsules such as the triangular and prolate structures (Fig. 3C-D). Non-icosahedral capsules have been seen experimentally in different T=7 viruses such as polyomavirus,⁴⁰ simian virus 40⁴¹ and particularly papillomavirus,⁴² whose empty viral capsids serve as cervical cancer vaccines.⁴³ Salunke *et al.*⁴⁰ observed both capsids that are larger and smaller than icosahedral capsids while Kanesashi *et al.*⁴¹ and Zhao *et al.*⁴² observed a range of tubular capsids. Although these T=7 viral capsids are experimentally assembled from 72 capsomers that are exclusively pentameric (without hexameric capsomers), they are still considered icosahedral capsids since the pentamers are arranged in local pentavalent and hexavalent environments that are quasi-equivalently symmetric; thus our model hexamers could be considered as pentamers in hexavalent environments.

Observation of small non-icosahedral capsules in T=9-19 systems

Large non-icosahedral capsules like those present in T=1-7 systems are not observed in T=9 and higher complexity systems. Instead, small non-icosahedral capsules similar to those present in the T=7 system (Fig. 3C-D) are observed in T=9, T=12 and T=13 systems (Set 1 in Fig. 4). In these systems, as more CPs are needed to enclose large T \geq 7 icosahedral capsids, partial capsids during assembly can afford to fold onto themselves prematurely to form capsules whose curvatures are similar to that of T=4 icosahedral capsids. For example, prolate capsules (Fig. 3D and 4) are formed by a partial capsid that contains 8 pentamers arranged symmetrically in two rows. This can be envisioned by discarding the four faces marked by the green arrow on the two-dimensional hexagonal surface in Fig. 1A and rolling the surface up along the direction indicated by the green arrow to fold into a three-dimensional structure. At each end, four hexamers are trapped in a ring forming a hole in the middle (Fig. 3E). The number of capsomers in these capsules observed in T \geq 7 systems can be generalized: each triangular capsule contains 3 tetrameric dislocations, 6 pentamers and 7(T-1) hexamers whereas each prolate capsule contains 2 tetrameric dislocations, 8 pentamers and 8(T-1) hexamers where T \geq 7.

Unlike the T=7 system, another set of triangular and prolate capsules emerges in T=9-13 systems (Set 2 in Fig. 4). In each T system, these capsules are smaller in curvature than those in Set 1. By examining the positions of pentamers in the middle of each prolate capsule one sees that the curvature of the Set 1 prolate capsule is provided by each row of four pentamers arranged in a *linear* pattern whereas the curvature of the Set 2 prolate capsule is provided by one row of four pentamers arranged in a *zigzag* pattern. This arrangement of pentamers results in a fewer number of faces and thus capsomers in Set 2 than that in Set 1. The number of capsomers in Set 2 capsules observed in T \geq 9 systems can be generalized: each triangular

capsule contains 6 pentamers and $5(T-1)+6\pm2$ hexamers and each prolate capsule contains 8 pentamers and $6(T-1)+4\pm2$ hexamers.

As the T number increases beyond T=13, there are other sets of triangular and prolate capsules whose curvatures are smaller than those observed in T=7-13 systems. For example, the T=16-19 systems exhibit another set of smaller closed non-icosahedral capsules (Set 3 in Fig. 5) and the T=19 system exhibits another set of yet smaller non-icosahedral capsules (Set 4 in Fig. 5). The curvatures of the Set 3 and Set 4 prolate capsules are provided by each row of three and two pentamers, respectively. The smallest curvature from these prolate capsules is similar to that of T=4 icosahedral capsids. In both T=16 and T=19 systems, it might be possible to assemble Set 1 capsules; however, none of those capsules emerge because the growing aggregates fold onto themselves to form smaller capsules earlier since they have as many capsomers that are needed to enclose a ring similar to the one from a T=4 icosahedral capsid. Consequently, self-assembly of larger structures such as empty icosahedral capsules is difficult; in fact, T=16 or T=19 icosahedral capsids were not observed in our simulations. This is true over a wide range of assembly conditions. At low protein concentrations, enclosed non-icosahedral capsules such those in Sets 2-4 are grown separately. At high protein concentrations, partially folded capsules are combined to form structures with multiple capsules of different types such as the joint capsules shown in Fig. 5.

The existence of polymorphic structures as observed in *in-vitro* experiments for many viruses and validated in this simulation-based study over a wide range of capsid systems (T=1-19) bridges our understanding of self-assembly of whole viruses *in vivo*, and other macromolecular assemblies such as micelle formation. On the one hand, *in vivo* assembly of each virus including the genomic content produces capsids of one size. On the other hand, assembly of micelles by amphiphilic co-polymers produces a polydisperse distribution of spherical and cylindrical clusters of different sizes.⁴⁴ Thus *in vitro* assembly of empty capsids is akin to micelle formation; however, there are significant differences. Firstly, the size distribution of micelles is continuous⁴⁴ whereas viral capsids/capsules vary by a discrete value because viral structures are completely enclosed and the tendency of capsid proteins to produce quasi-equivalent conformations that arrange themselves at a regularly spaced distance. Secondly, micelles are not as ordered as viral structures which adopt the highest symmetry for a closed polyhedron. Therefore, capsid assembly is entropically less favorable since each incoming CP subunit has to position itself at the right location without much structural flexibility. Nevertheless, it is remarkable that the self-assembly of empty capsids, without any help from auxiliary machinery, is still possible.

Observation of empty icosahedral capsids in T=1-13 systems

Self-assembly of icosahedral capsids at optimal conditions of 43.5 M and 308K in different capsid systems ranging from T=1 to T=13 is observed to share the same kinetic mechanism involving elementary kinetics via the sequential, reversible addition of predominately capsomers. Here, self-assembly is a nucleated process involving low order aggregates that exist predominantly early in the simulations. Once formed, such intermediates are very 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 capsomers. This finding is similarly observed in light scattering and size exclusion chromatography experiments by Zlotnick and coworkers^{45,46} and in previous simulations using simplified models^{10,12,13,18}

The high icosahedral capsid yield in T=1-7 systems indicates that although the role of scaffold proteins and genomic content play an important role in the self-assembly of structure in many viruses, it is possible to assemble them using only coat proteins. For assembly of larger capsids, self-assembly in the absence of any supporting materials results in low yields (< % 10)

indicating that the role of scaffold proteins and genomic content becomes more significant, as they serve to stabilize the growing structures and prevent them from folding onto themselves. This may help explain the lack of experimental evidence for successfully assembling empty capsids of $T \geq 9$ *in vitro* without the help of supporting materials.

Conclusion

In summary, this study provides a theoretical framework for understanding structure polymorphism in a wide range of capsid systems. By combining simple models with discontinuous molecular dynamics,²⁴ we have been able to simulate the spontaneous formation of supramolecular structures of $T=1-19$ viral capsids initiated from random configurations of proteins in the forms of either pentameric or hexameric capsomers. Our simulations elucidate the kinetic mechanism of self-assembly of icosahedral capsids and a variety of non-icosahedral capsules with similar energetics and stability. These non-icosahedral capsules share a morphology similar to those observed in experiments on many virus systems. Our findings coupled with the results from our previous simulation studies, in which CPs are started as either monomeric, dimeric or trimeric, demonstrate that structural polymorphism in the endpoint structures is an inherent property of CPs and arises from condition-dependent kinetic mechanisms, independent of the capsid T number and the elemental mechanism of capsid assembly. However, there are two distinctive classes of non-icosahedral capsules: those that are larger than icosahedral capsids in $T \leq 7$ systems and capsules smaller than icosahedral capsids in $T \geq 7$ systems. The latter class includes multiple sets of capsules; the exact number of capsule sets expands as T increases as long as the curvature of these capsules is at least similar to that of $T=4$ icosahedral capsids. The composition of this underlying phase space shifts to non-icosahedral capsules in $T \geq 9$ systems suggesting that the presence of other supporting materials is needed for the self-assembly of icosahedral capsids. Based on the results from this study, future research should focus on formulating the means to control the self-assembly of different structures by either stabilizing the assembly intermediates (e.g., spatial or temporal control) or accounting for the role of the genomic content and/or other auxiliary components such as scaffold proteins in the self-assembly of viral capsids and full viruses.

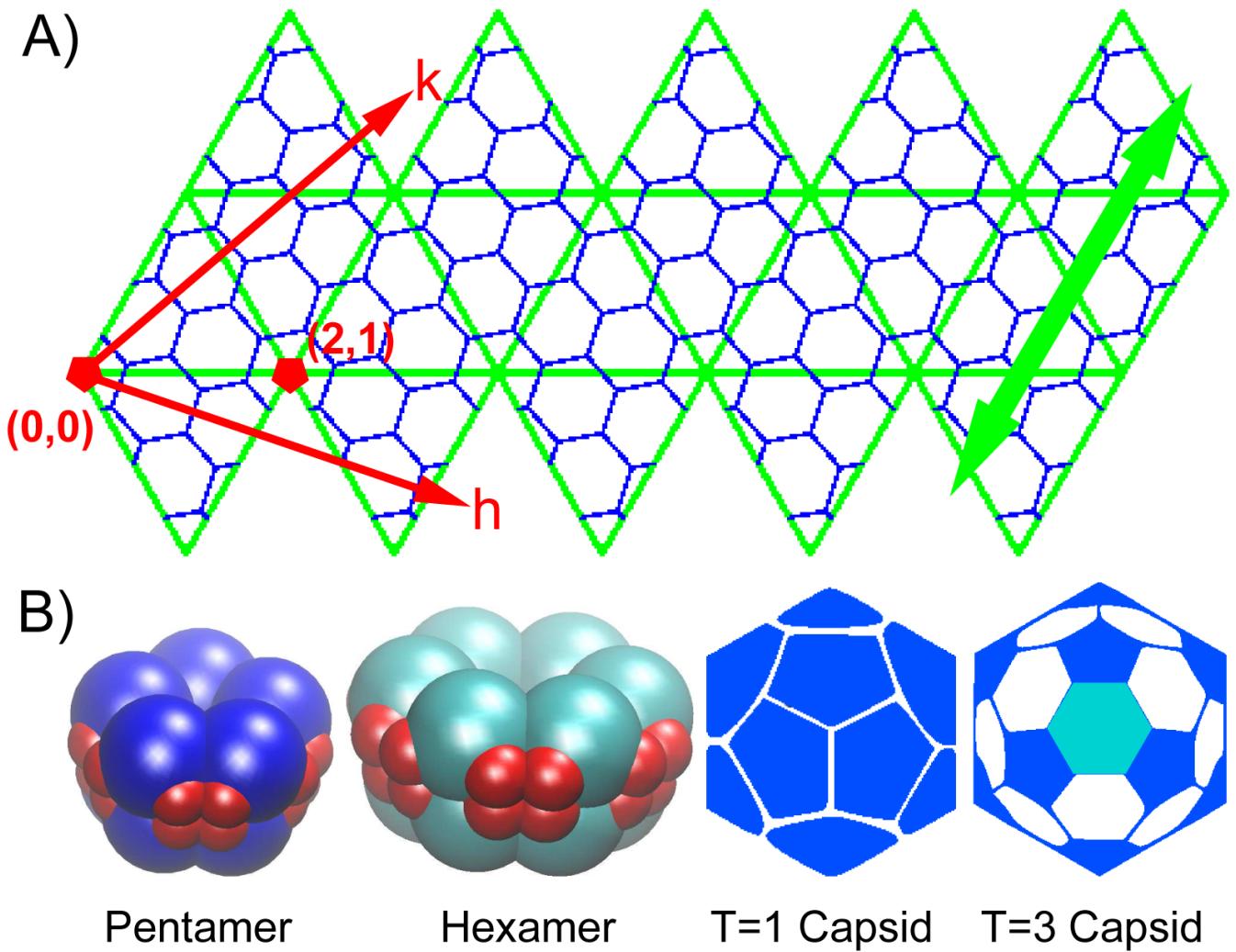
Acknowledgments

We would like to thank Vijay S. Reddy for helpful discussions and acknowledge NIH (RR12255) and NSF (PHY0216576) for financial support.

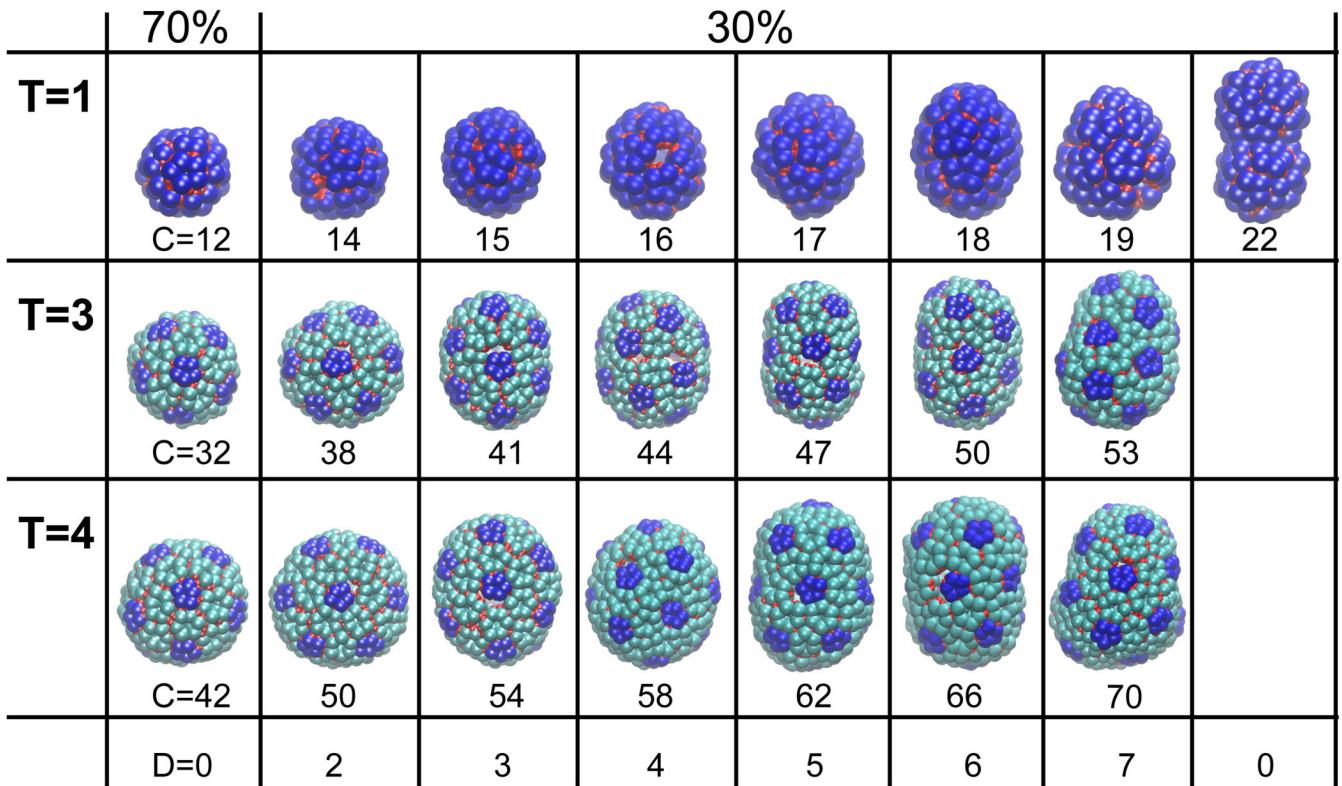
References

- (1). Crick FH, Watson JD. Nature 1956;177:473. [PubMed: 13309339]
- (2). Caspar DLD, Klug A. Cold Spring Harb. Symp Quant. Biol 1962;27:1. [PubMed: 14019094]
- (3). Speir JA, Munshi S, Wang G, Baker TS, Johnson JE. Structure 1995;3:63. [PubMed: 7743132]
- (4). Harrison SC. Curr Opin Struct Biol 2001;11:195. [PubMed: 11297927]
- (5). Fox JM, Wang G, Speir JA, Olson NH, Johnson JE, Baker TS, Young MJ. Virology 1998;244:212. [PubMed: 9581792]
- (6). Bancroft JB, Bracker CE, Wagner GW. Virology 1969;38:324. [PubMed: 5784855]
- (7). Dong XF, Natarajan P, Tihova M, Johnson JE, Schneemann A. J. Virol 1998;72:6024. [PubMed: 9621065]
- (8). Douglas T, Young M. Science 2006;312:873. [PubMed: 16690856]
- (9). Rapaport DC. Phys Rev E Stat Nonlin Soft Matter Phys 2004;70:051905. [PubMed: 15600654]
- (10). Zhang T, Schwartz R. Biophys J 2006;90:57. [PubMed: 16214864]
- (11). Hemberg M, Yaliraki SN, Barahona M. Biophys J 2006;90:3029. [PubMed: 16473916]
- (12). Hagan MF, Chandler D. Biophys J 2006;91:42. [PubMed: 16565055]
- (13). Nguyen HD, Reddy VS, Brooks CL III. Nano Lett 2007;7:338. [PubMed: 17297998]

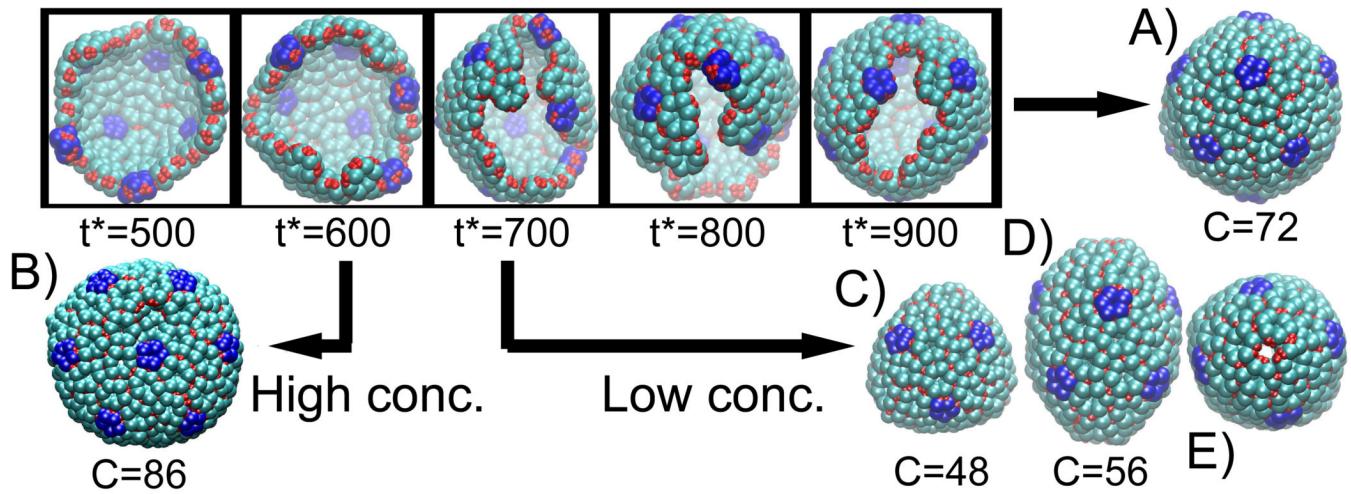
- (14). Bruinsma RF, Gelbart WM, Reguera D, Rudnick J, Zandi R. *Phys Rev Lett* 2003;90:248101. [PubMed: 12857229]
- (15). Nguyen TT, Bruinsma RF, Gelbart WM. *Phys Rev Lett* 2006;96:078102. [PubMed: 16606144]
- (16). Keef T, Taormina A, Twarock R. *Journal of Physics-Condensed Matter* 2006;18:S375.
- (17). Chen T, Glotzer SC. *Phys Rev E Stat Nonlin Soft Matter Phys* 2007;75:051504. [PubMed: 17677070]
- (18). Nguyen HD, Reddy VS, Brooks CL III. In submission. 2008
- (19). Mannige R, Brooks CL III. *Phys Rev E* 2008;77:051902.
- (20). Xie Z, Hendrix RW. *J Mol Biol* 1995;253:74. [PubMed: 7473718]
- (21). Zandi R, Reguera D, Bruinsma RF, Gelbart WM, Rudnick J. *Proc Natl Acad Sci U S A* 2004;101:15556. [PubMed: 15486087]
- (22). Chen T, Zhang Z, Glotzer SC. *Proc Natl Acad Sci U S A* 2007;104:717–22. [PubMed: 17215354]
- (23). Kerner R. *Computational and Mathematical Methods in Medicine* 2008;9:175.
- (24). Alder BJ, Wainwright T. *J. Chem. Phys* 1959;31:459.
- (25). 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. [PubMed: 16381893]
- (26). Ceres P, Zlotnick A. *Biochemistry* 2002;41:11525. [PubMed: 12269796]
- (27). Johnson JM, Tang J, Nyame Y, Willits D, Young MJ, Zlotnick A. *Nano Lett* 2005;4:765. [PubMed: 15826125]
- (28). Tong LA, Rossmann MG. *Acta Crystallogr A* 1990;46:783. [PubMed: 2174243]
- (29). Sorger PK, Stockley PG, Harrison SC. *J Mol Biol* 1986;191:639. [PubMed: 3806677]
- (30). Earnshaw W, King J. *J. Mol. Biol* 1978;126:721. [PubMed: 370407]
- (31). Wikoff WR, Johnson JE. *Curr Biol* 1999;9:R296. [PubMed: 10226016]
- (32). Adolph KW, Butler PJ. *J Mol Biol* 1974;88:327. [PubMed: 4452998]
- (33). Fu CY, Morais MC, Battisti AJ, Rossmann MG, Prevelige J. *J Mol Biol* 2007;366:1161. [PubMed: 17198713]P. E.
- (34). Cusack S, Oostergetel GT, Krijgsman PC, Mellema JE. *J Mol Biol* 1983;171:139. [PubMed: 6655690]
- (35). Ganser-Pornillos BK, von Schwedler UK, Stray KM, Aiken C, Sundquist WI. *J Virol* 2004;78:2545. [PubMed: 14963157]
- (36). Rossmann MG. *Virology* 1984;134:1. [PubMed: 6710869]
- (37). Zhao X, Fox JM, Olson NH, Baker TS, Young MJ. *Virology* 1995;207:486. [PubMed: 7886952]
- (38). Francki RIB, Hatta T, Boccardo G, Randles JW. *Virology* 1980;101:233. [PubMed: 18631638]
- (39). Zhang W, Olson NH, Baker TS, Faulkner L, Agbandje-McKenna M, Boulton MI, Davies JW, McKenna R. *Virology* 2001;279:471. [PubMed: 11162803]
- (40). Salunke DM, Caspar DL, Garcea RL. *Biophys J* 1989;56:887. [PubMed: 2557933]
- (41). Kanesashi SN, Ishizu K, Kawano MA, Han SI, Tomita S, Watanabe H, Kataoka K, Handa H. *J Gen Virol* 2003;84:1899. [PubMed: 12810885]
- (42). Zhao Q, Guo HH, Wang Y, Washabaugh MW, Sitrin RD. *J Virol Methods* 2005;127:133. [PubMed: 15894387]
- (43). Koutsky LA, Ault KA, Wheeler CM, Brown DR, Barr E, Alvarez FB, Chiacchierini LM, Jansen KU. *N. Engl. J. Med* 2002;347:1645. [PubMed: 12444178]
- (44). Jones, MN.; Chapman, D. *Micelles, monolayers and biomembranes*. Wiley-Liss; New York, NY: 1995.
- (45). Zlotnick A, Johnson JM, Wingfield PW, Stahl SJ, Endres D. *Biochemistry* 1999;38:14644. [PubMed: 10545189]
- (46). Endres D, Zlotnick A. *Biophys J* 2002;83:1217. [PubMed: 12124301]
- (47). Johnson JE, Speir JA. *J Mol Biol* 1997;269:665. [PubMed: 9223631]
- (48). Reddy VS, Natarajan P, Okerberg B, Li K, Damodaran KV, Morton RT, Brooks CL III, Johnson JE. *J. Virol* 2001;75:11943. [PubMed: 11711584]

**Figure 1.**

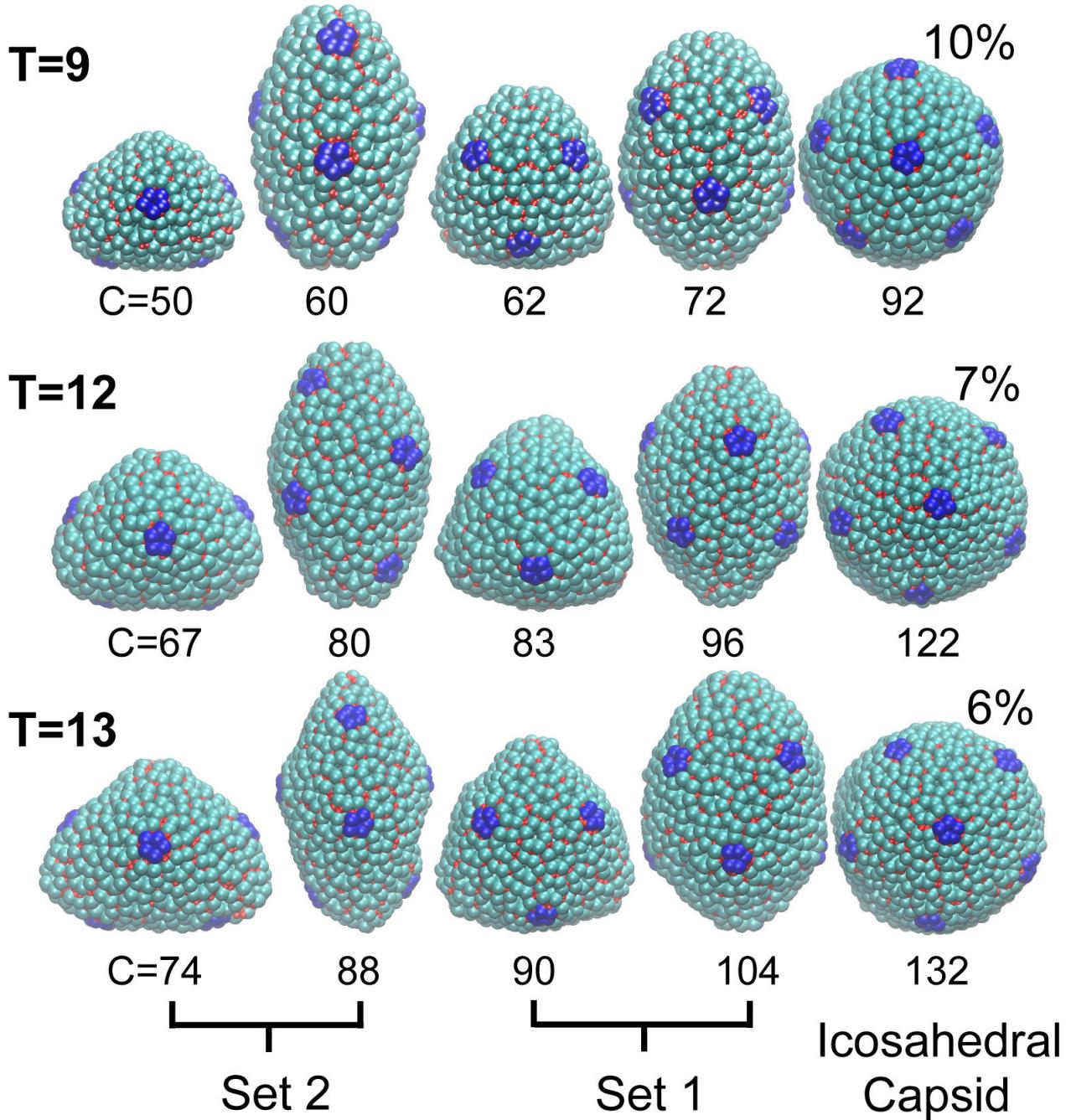
A) Geometric principles for generating icosahedral quasi-equivalent surface lattices: on a planar hexagonal surface, 12 curvature-inducing pentamers are inserted at defined positions along the h and k axis. An icosahedron shown as an example in two dimensions with 20 identical triangular faces is identified as a $T=h^2 + h * k + k^2 = 7$ capsid ($h=2$ and $k=1$); the other enantiomeric $T=7$ capsid ($h=1$ and $k=2$) is also considered.⁴⁷ (Template taken from Viperdb⁴⁸). B) Simple models representing each fivefold pentameric capsomer as a pentagonal structure and each sixfold hexameric capsomer as a hexagonal structure are utilized. The large spheres are repulsive providing volume exclusion. The small red spheres are attractive providing inter-capsomer interactions if they on similar middle layers and belong to the interfaces that are specified according to the T capsid system. For example, the T=1 system, without hexamers, requires pentamers to be attractive to one another at any interface whereas the T=3 system requires each hexamer to be attractive to any pentamer at every other interfaces and to other hexamers at alternating interfaces. The model permits formation of “native” interfaces, if the two capsomers are aligned at the correct angle, with four attractive interactions; “non-native” interfaces are formed if two capsomers are locked at a wrong angle, with fewer attractive interactions.

**Figure 2.**

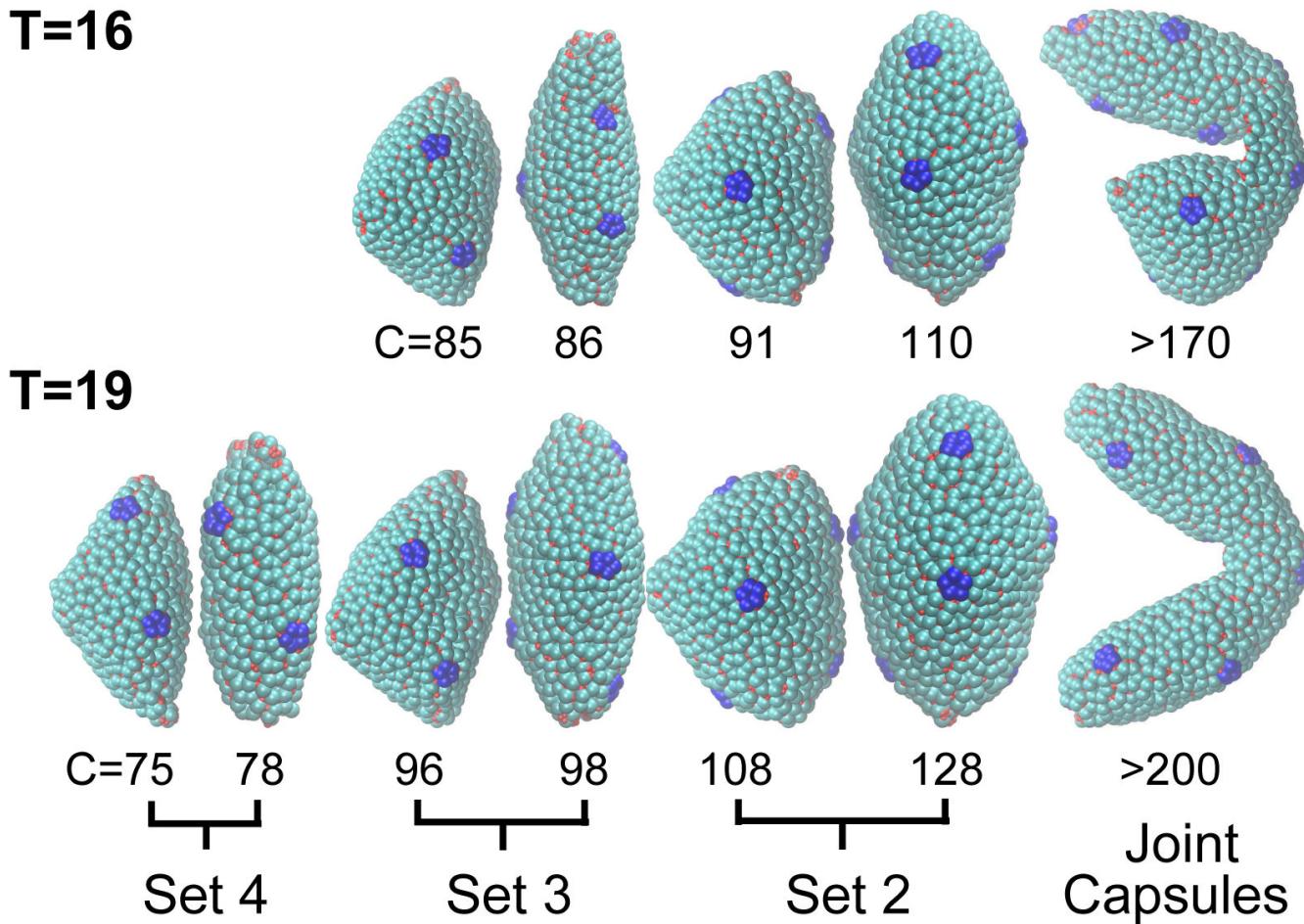
The population distribution for supramolecular structures obtained from $T=1$, $T=3$ and $T=4$ systems at $43.5 \mu\text{M}$ and 290K as a function of the number of hexameric dislocations D : complete icosahedral capsids; non-icosahedral yet highly symmetric enclosed capsules: oblate, angular, twisted, tubular, oblong, conical and geminate. The icosahedral capsid yield is about 70% in the three systems while the rest belong to non-icosahedral capsules whose sizes are inversely proportional to their yields.

**Figure 3.**

Snapshots of aggregate particles that grow into an icosahedral T=7 capsid (A) over time (t^* in reduced units) at an optimal condition of $43.5 \mu\text{M}$ and 308K . At higher protein concentrations, off-pathway mechanism involving the collapse event by sizable intermediates results in the formation of non-icosahedral yet closed capsules such as oblate structures (B) that are larger than icosahedral capsids. At lower protein concentrations, off-pathway mechanism involving prematurely enclosed intermediate results in the formation of non-icosahedral yet closed capsules such as triangular (C) and prolate (D) structures that are smaller than icosahedral capsids. These capsules contain tetrameric dislocations at which four hexamers are kinetically trapped in a ring forming a hole in the middle (E).

**Figure 4.**

The population distribution for supramolecular structures obtained from $T=9$, $T=12$ and $T=13$ systems at $43.5 \mu\text{M}$ and 290K as a function of the number of capsomers C (either pentamers or hexamers) showing icosahedral capsids and non-icosahedral yet highly symmetric enclosed capsules that are smaller than icosahedral capsids: Set 1, which also exists in the $T=7$ system (Fig. 3C-D), consists of triangular capsules with 6 pentamers and $7(T-1)$ hexamers and prolate capsules with 8 pentamers and $8(T-1)$ hexamers; Set 2 consists of triangular capsules with 6 pentamers and $5(T-1)+6\pm 2$ hexamers and prolate capsules with 8 pentamers and $6(T-1)+4\pm 2$ hexamers.

**Figure 5.**

Represented enclosed supramolecular structures obtained from T=16 and T=19 systems over a wide range of assembly conditions: non-icosahedral yet highly symmetric enclosed capsules in different sizes and joint structures containing multiple capsules of different types. The Set 2 capsules are also observed in T=9-13 systems; additional sets of non-icosahedral capsules emerge as T increases from T=13 to T=16 and then T=19. All of these single non-icosahedral capsules are smaller than their respective icosahedral capsids (C=162 and 192 capsomers for T=16 and 19, respectively), which are too large to be assembled since each growing aggregate tends to fold onto itself to form non-icosahedral capsules of a variety of sizes.