NIH Public Access
Author Manuscript
J Comput Chem. Author manuscript; available in PMC 2012 May 15.

Published in final edited form as:
J Comput Chem. 2009 February ; 30(3): 423–437. doi:10.1002/jcc.21071.

Molecular Dynamics/Order Parameter eXtrapolation (MD/OPX) for Bionanosystem Simulations

Yinglong Miao and **Peter J. Ortoleva**
Center for Cell and Virus Theory, Department of Chemistry, Indiana University, Bloomington, Indiana 47405, USA

Abstract

A multiscale approach, Molecular Dynamics/Order Parameter eXtrapolation (MD/OPX), to the all-atom simulation of large bionanosystems is presented. The approach starts with the introduction of a set of order parameters (OPs) automatically generated with orthogonal polynomials to characterize the nanoscale features of bionanosystems. The OPs are shown to evolve slowly via Newton's equations and the all-atom multiscale analysis (AMA) developed earlier[1] demonstrates the existence of their stochastic dynamics, which serve as the justification for our MD/OPX approach. In MD/OPX, a short MD run estimates the rate of change of the OPs, which is then used to extrapolate the state of the system over time that is much longer than the 10^{-14} second timescale of fast atomic vibrations and collisions. The approach is implemented in NAMD and demonstrated on cowpea chlorotic mottle virus (CCMV) capsid structural transitions (STs). It greatly accelerates the MD code and its underlying all-atom description of the nanosystems enables the use of a universal inter-atomic force field, avoiding recalibration with each new application as needed for coarse-grained models.

Keywords

bionanosystems; CCMV; structural transitions; order parameters; all-atom multiscale analysis; MD/OPX

I Introduction

Bionanostructures are key elements of many systems of interest in the pure and applied life sciences. Natural examples include viruses, ribosomes and other sub-cellular features. Engineered bionanosystems include vaccines (i.e. diminished viruses), nanoparticles functionalized for medical imaging and thermal cancer treatments, and nanocapsules for targeted delivery of therapeutic agents. These systems typically have supra-million atoms in size. They reside at the interface between the highly atomic fluctuating and the near-macroscopic deterministic regimes and their atomic-scale detail is vital to their function. As a consequence, they display dual atomistic/macroscopic behaviors on multiple spatial and temporal scales, which must be understood simultaneously.

Our objective is to develop a methodology to simulate these systems that builds in a strategy designed to address the special characteristics of the multiple scale processes and the cross-talk among as follows. Atomic-scale fluctuations of nanosystems provide the entropy involved in the thermal average forces driving their nanoscale dynamics. Conversely, their overall transformations modify the statistics of the atomic-scale fluctuations.

812-855-2717, ortoleva@indiana.eduhttp://sysbio.indiana.edu.

Viruses present an excellent example of this inter-scale cross-talk. They display a variety of structural dynamical phenomena and hence are used to motivate this study. Many viruses undergo STs, i.e. large conformational changes in their architecture, in response to changes in their microenvironment conditions. The later include temperature, pH, ionic strength, and cation concentrations, as well as contact with a host cell membrane. For example, native cowpea chlorotic mottle virus (CCMV) capsid swells dramatically as pH is increased from 5.0 to 7.5 in the absence of divalent cations[2–6]. Nudaurelia capensis ω virus (NωV)[7–10] and HK97 bacteriophage[11,12] undergo large structural changes during capsid maturation. Poliovirus undergoes irreversible STs upon maturation or receptor-mediated cell entry[13–17]. Such STs are believed to play significant roles in viral life-circle events, such as virus attachment to cell membranes and genome release from the viral capsid into a host cell. As characterized quantitatively via electron cryo-microscopy (cryoEM) and X-ray crystallography, these STs are typically accomplished through a series of distinct steps involving intact translation and rotation of protein structural units (e.g. protomers, pentameric and hexameric capsomers) and those in poliovirus and NωV are triggered by irreversible bond-cleavage of capsid proteins.

In principle, many aspects of viral STs, assembly and disassembly can be understood in terms of the dynamics of a set of atoms evolving classically through an interatomic force field. The benefit of such an approach is that it enables a general, parameter-free predictive nanosystem model. The challenge is to implement this approach as a practical computational algorithm. A central objective of the approach presented here is to capture atomic-scale detail and yet simulate whole-virus dynamics.

Theoretical approaches have been developed to simulate viral processes. These include coarse-grained models[18,19], symmetry-constrained models[20], computational molecular dynamics (MD)[20–33] and normal mode analysis[34–37] as reviewed earlier[1,38]. However, coarse-grained models miss the atomic detail needed for interpreting drug and cell surface receptor interactions, and do not allow the universality afforded by an interatomic force field. Normal mode analysis is based on single-well (harmonic) potential approximations. It provides some insights into viral STs, but cannot be used to simulate STs, and the calculated vibrational normal modes can not describe the diffusive (friction-dominated) motion that dominates slow viral processes; thus it cannot address the very local and highly nonlinear nature of the drug-virus interaction, or the initiation and propagation of a structural transition across a virus. Direct MD is limited to biologically uninteresting timescales. For example, NAMD[39], a high-performance parallel MD code, has been used to simulate a complete satellite tobacco mosaic virus (STMV). The simulation took 50 days on a SGI 1,024 processor Altrix system using 256 processors and 128 GB of memory, but only captured a physical time of 55 nanoseconds (ns) [21]. However, viral STs take milliseconds (ms) to minutes or longer[8]. With such an approach it would take about 2,500 years to obtain a physical result. Clearly a speed-up of at least a factor of 10,000 over MD is needed to make all-atom simulation a viable approach. In summary, previous approaches do not simultaneously capture atomic-scale detail and whole-virus, long-time behaviors in a nonsymmetry-constrained fashion, or account for the highly nonlinear and the often symmetry-broken nature of many key bionanosystem phenomena. The challenge remains to develop methods that preserve the atomic/nanoscale cross-talk underlying the dynamics of bionanosystems.

Here we present MD/OPX, an approach for all-atom simulation of large bionanosystems with the potential to overcome the shortcomings of the methods noted above. MD/OPX is based on our all-atom multiscale theory of nanosystem dynamics[1,38]. It integrates the following elements: (1) automated construction of order parameters (OPs) characterizing the nanoscale structural features of nanosystems, (2) a demonstration based on Newton’s equations that prove the OPs evolve slowly relative to the 10^{-14} second timescale of fast atomic vibrations and collisions, (3) rigorous multiscale techniques that can be used to derive a stochastic (Fokker-Planck or Smoluchowski) equation and equivalent Langevin equations of the dynamics of these OPs starting from the Liouville equation, (4) the use of (1) to (3) to justify the suggestion of the equation free multiscale (EFM) analysis approach developed by Kevrekidis *et al.* [40], that short bursts of MD simulations can be used to extrapolate coarse variables (i.e. OPs) over large time intervals and thus advance the nanostructure over long time, and (5) an implementation of (1) to (4) optimized and stabilized via an adaptive dual timestep procedure and microscopic configuration-corrected extrapolation using our automatically generated OPs. The MD/OPX approach is essentially equivalent to solving the Langevin equations, except that the friction coefficients and thermal-average forces are evaluated implicitly.

In the following, we generate general OPs automatically for bionanosystems in Sect. II, present MD/OPX as an approach to solving the stochastic dynamics of these OPs implicitly in Sect. III, and demonstrate MD/OPX on CCMV capsid STs in Sect. IV. Conclusions are drawn in Sect. V.

II General Order Parameters for Bionanosystems

The order parameters (OPs) for bionanosystems considered in this study have the following properties:

- capture overall nanoscale features of the nanosystems;
- vary much more slowly in time than the 10^{-14} second timescale of fast atomic vibrations and collisions;
- can be expressed in terms of the atomic positions and momenta; and
- can be generated via a readily implemented, general procedure capturing complex features of interest in virology and cell physiology.

A brief survey of the types of phenomena these OPs are intended to capture sets the stage for our formulation.

A nanoparticle in an aqueous medium has internal dynamics, e.g. the STs of a viral capsid or genome. As the nanoparticle migrates across the system, it often has a closely associated layer of water molecules. The motion of the nanoparticle may excite hydrodynamic modes in the host fluid. Multiple nanoparticles, viruses, nanocapsules with therapeutic payloads, and other elements may interact – constituting a complex, composite of nanoscale subunits that may repel each other, bind, coalesce or self-assemble into nanomaterials. We seek a method to construct OPs that can be programmed to capture the salient features of such systems and phenomena. Our algorithm is as follows.

Consider a system embedded in a box (a cube for simplicity of presentation) of volume L^3. Basis functions $u_k(x)$ (e.g., polynomials or harmonic functions) labeled with integer index k are introduced such that

$$\int_{-L/2}^{L/2} dx u_k(x) u_{k'}(x) = \delta_{kk'} \tag{II.1}$$

for Kronecker delta $\delta_{kk'}$. Composite functions $U_{\underline{k}}(\vec{s})$ are defined such that

$$U_{\underline{k}}(\vec{s}) = u_{k_1}(x) u_{k_2}(y) u_{k_3}(z), \tag{II.2}$$

where $\vec{s} = (x,y,z)$ for box centered at $\vec{s} = \vec{0}$. According to the space-warping method[41], a nanostructure embedded in a space ($\vec{s}$ here) is considered to be a deformation of a reference space $\vec{s}^{\,0}$. The deformation of space is used to introduce OPs via

$$\vec{s} = \sum_{\underline{k}} U_{\underline{k}}(\vec{s}^{\,0})\vec{\Psi}_{\underline{k}}. \tag{II.3}$$

As the $\vec{\Psi}_{\underline{k}}$ change, $\vec{s}$-space is deformed, and so is the nanosystem embedded in it. The $\vec{\Psi}_{\underline{k}}$ constitute a set of vector OPs that serve as the starting point of our all-atom multiscale approach provided that they can be related to the atomic configuration of the nanosystem and can be shown to evolve slowly.

Each atom in the system is moved via the above deformation by evolving $\vec{\Psi}_{\underline{k}}$. However, given a finite truncation of the $\underline{k}$ sum in (II.3), there will be some residual displacement of individual atoms that is not accounted for. Denoting this residual for atom i as $\vec{\sigma}_i$, we write

$$\vec{s}_i = \sum_{\underline{k}} U_{\underline{k}}(\vec{s}_i^{\,0})\vec{\Psi}_{\underline{k}} + \vec{\sigma}_i. \tag{II.4}$$

This gives $\vec{s}_i$ in the deformed state for the instantaneous values of the $\vec{\Psi}_{\underline{k}}$ and $\vec{\sigma}_i$, and the reference configuration $\vec{s}_i^{\,0}$. If a sufficient number of terms are retained in the sum and the basis functions are chosen properly, then the $\vec{\sigma}_i$ will be small. The exception to this theme is for cases where the inter-diffusion of molecules is a key part of the phenomenon of interest, e.g. viscous drug molecules in an aqueous medium. This follows because there is an angstrom distortion of space needed to keep track of the motion of each small molecule relative to its neighbors. In this case we expect that the $\vec{\sigma}_i$ could grow as square root of time, i.e. to display typical random walk behaviors. However, even this case could be addressed as the system configuration is extremely stray so that the reference configuration $\vec{s}_i^{\,0}$ $(i=1,2,\cdots)$ could be regularly updated at a type of "piecewise continuation" in time by monitoring the $\vec{\sigma}_i$ in the course of a simulation.

To evolve $\vec{\Psi}_{\underline{k}}$ via Newtonian mechanics, and thereby start our multiscale analysis, $\vec{\Psi}_{\underline{k}}$ must be expressed in terms of the $\vec{s}_i$. Let m_i be the mass of atom i and m be the mass of a typical atom. Multiplying (II.4) by $m_i U_{\underline{q}}\left(\vec{s}_i^{\,0}\right)/m$ and summing over the N atoms ($i = 1,2,\cdots N$) in the system yieleds

$$\sum_{\underline{k}} B_{\underline{qk}}\vec{\Psi}_{\underline{k}} = \frac{L^3}{N}\sum_{i=1}^{N}\frac{m_i}{m}U_{\underline{q}}\left(\vec{s}_i^{\,0}\right)\vec{s}_i,\; B_{\underline{qk}} = \frac{L^3}{N}\sum_{i=1}^{N}\frac{m_i}{m}U_{\underline{q}}\left(\vec{s}_i^{\,0}\right)U_{\underline{k}}\left(\vec{s}_i^{\,0}\right). \tag{II.5}$$

The $\vec{\sigma}_i$ contribution is neglected in arriving at this definition of $\vec{\Psi}_{\underline{k}}$ as $\vec{\sigma}_i$ fluctuates with i and hence with space, while the basis functions that capture overall nanostructural features vary smoothly by design, i.e., capture overall features such as nanoparticle position, orientation, size and shape. Thus, (II.5) is not an approximation; rather, the above discussion is a way to argue for the definition (II.5) of OPs that captures coherent behaviors of a nanosystem. With this definition of $\vec{\Psi}_{\underline{k}}$, (II.5) is an exact relationship, since the $\vec{\sigma}_i$ correct errors in the displaced atomic positions over-and-above the coherent contribution from the $\vec{\Psi}_{\underline{k}}$ sum.

Our approach has several conceptual and technical advantages. In the above we include all atoms in the system, e.g., those in a nanoparticle and those in its microenvironment. This allows the formulation to capture the boundary layer of water that tends to move with a nanoparticle due to the viscous nature of water at the nanoscale. Thus the method captures hydrodynamic modes in the host medium and layers of water bound to the nanoparticle. This formulation also allows for the description of multiple nanoparticles, viral capsids composed of protein capsomers, or a virus interacting with a host cell membrane. The orthogonality of the basis functions (II.1) implies that the matrix B is nearly diagonal and the OPs can easily be computed numerically in terms of the atomic positions. Especially when most of the space in the system is occupied with atoms, the i sum is essentially a Monte Carlo integration. The orthogonality of the normailized basis functions implies that $B_{\underline{qk}} \approx \delta_{\underline{qk}}$ and (II.5) can be approximated as

$$\vec{\Psi}_{\underline{q}} \approx \frac{L^3}{N} \sum_{i=1}^{N} \frac{m_i}{m} U_{\underline{q}}\left(\vec{s}_i^{\,0}\right) \vec{s}_i. \qquad \text{(II.6)}$$

In our implementation of the method, we provide both options to solve a linear system of equations in (II.5) for $\vec{\Psi}_{\underline{k}}$ or compute them directly via (II.6). Inclusion of m_i in the above expressions gives the OPs the character of generalized center-of-mass (CM) variables. Thus, if $U_{\underline{k}}$ is constant then $\vec{\Psi}_{\underline{k}}$ is proportional to the CM.

In the above formulation, we did not make any assumptions about the structure of the nanosystem. Thus there may be one nanoparticle in an aqueous host, multiple nanoparticles, a virus and a cell membrane, or a lipsome with its therapeutic payload. The ability to address composite nanosystems is a great benefit of the present formulation.

Newton's equations and statistical considerations can be used to argue that the above OPs $\vec{\Psi}_{\underline{k}}$ evolve slowly. As $d\vec{s}_i/dt = \vec{p}_i/m_i$ for momentum $\vec{p}_i$ of atom i, one has

$$\begin{gathered} \frac{d\vec{\Psi}_{\underline{k}}}{dt} = \frac{\vec{\Pi}_{\underline{k}}}{Nm}, \\ \vec{\Pi}_{\underline{k}} \approx L^3 \sum_{i=1}^{N} U_{\underline{q}}\left(\vec{s}_i^{\,0}\right) \vec{p}_i. \end{gathered} \qquad \text{(II.7)}$$

While $\vec{\Psi}_{\underline{k}}$ has a sum of N atoms many of which have similar directions due to the smooth variation in $U_{\underline{k}}$ with $\vec{s}_i^{\,0}$, the momenta have fluctuating direction and tend to cancel near equilibrium. Hence thermal average of $\vec{\Pi}_{\underline{k}}$ is small and hence $\vec{\Psi}_{\underline{k}}$ tend to evolve slowly and the ratio of the characteristic time of $\vec{\Psi}_{\underline{k}}$ to that of fast atomic vibrations and collisions should be on the order of the number of atoms in the system, i.e., O(N). This argument has been validated via simulations as will be discussed in Sect. IV.

III Molecular Dynamics/Order Parameter eXtrapolation (MD/OPX)

Newton's equations show that if $U_{\underline{k}}$ vary smoothly on a nanometer scale, then the OPs $\vec{\Psi}_{\underline{k}}$ evolve on a timescale longer than the 10^{-14} second timescale of fast atomic vibrations and collisions. In this case, our AMA[1,38,42] demonstrates the existence of a stochastic (Fokker-Planck or Smoluchowski) equation and equivalent Langevin equations for the dynamics of these Ops $\vec{\Psi}_{\underline{k}}$. In this section, we will present MD/OPX, a computational approach to solving the Langevin equations implicitly without the need to construct the thermal-average forces and friction coefficients.

Since an MD simulation generates the evolution of the all-atom configuration of a nanostructure, it yields our OPs via (II.5). In a small time interval δt (e.g., 100 MD timesteps), an MD code output can be used to generate $\vec{\Psi}_{\underline{k}}(t+\delta t)$. Then, the change $\vec{\Psi}_{\underline{k}}(t+\delta t) - \vec{\Psi}_{\underline{k}}(t)$ divided by δt represents the deterministic part of the Langevin dynamics. With this, one approximates $\vec{\Psi}_{\underline{k}}(t+\Delta t)$ for a timestep Δt via $\vec{\Psi}_{\underline{k}}(t+\Delta t) \approx \vec{\Psi}_{\underline{k}}(t) + \Delta t\,[\vec{\Psi}_{\underline{k}}(t+\delta t)$

$-\Psi_{\underline{k}}(t)]/\delta t$, where $\Delta t \gg \delta t$ but smaller than the characteristic time for OP dynamics. In this way, one may advance $\Psi_{\underline{k}}$ via a sequence of dual (δt, Δt) cycles.

To implement this dual timestep algorithm, we use (II.4) to reconstruct the all-atom configuration after the order parameter advancement to $t + \Delta t$, taking the residual σ_i to be its value at $t + \delta t$. We justify the use of $\sigma_i(t + \delta t)$ to approximate $\sigma_i(t + \Delta t)$ because we find that (1) with a sufficient number of basis functions $U_{\underline{k}}$ chosen the σ_i are small and (2) various ways (e.g. energy minimization and short MD run) can be used to "anneal" them to be consistent with states of nanostructures likely to appear under biological conditions.

In the present implementation of MD/OPX, we use NAMD as our platform. A Fortran program was written to read NAMD output all-atom structures, calculate the resultant OPs by solving (II.5), extrapolate the OPs for Δt, generate an atomic configuration at $t + \Delta t$ with the extrapolated OPs and put the result all-atom structure back into NAMD to start the next (δt, Δt) cycle. The flowchart in Fig. 1(a) shows this MD/OPX work flow. Starting from a PDB structure, an initial minimization and an initial MD run are applied before MD/OPX cycling commences.

A variety of steps were taken to optimize and stabilize the MD/OPX computations. First, the OPX over Δt deforms all space continuously. This can lead to unphysical atomic structures, such as stretched bonds, unreasonable bond angles and dihedrals, close/overlapping atoms, etc. Thus we perform an energy minimization and a short MD run to anneal these unphysical structures after obtaining the all-atom structure that is generated from extrapolated OPs. Next we implement an adaptive cycling. Simulation parameters including δt, Δt, the amount of energy minimization steps and the length of the short MD run (δt_0) for structure annealing are adjusted dynamically during program running to optimize the balance between accuracy and CPU speed. Total forces on atoms of the simulated structure are loaded and atomic accelerations are calculated as indicators of the unphysical structures after OPX. Specifically, the average atomic acceleration of the evolving structure after OPX and the maximum atomic acceleration after energy minimization are used as the indicators for on-the-fly parameter control. If the indicators are within the chosen boundaries, our code increments $\Delta t/\delta t$ ratio by a factor of 1.1 to accelerate the simulation. If they are beyond the critical values, the $\Delta t/\delta t$ ratio is decreased by a factor of 0.5. The difference between the acceleration and deceleration feature was found to stabilize and optimize the overall algorithm. The objective of our adaptive procedure is to dynamically seek the optimum – i.e., to minimize the total CPU time for a given biological time by varying the run parameters. The final flowchart is shown in Fig. 1b.

Our MD/OPX implementation greatly enhances the utility of standard MD codes for large bionanosystem applications. Options provided by MD/OPX include: (1) adaptive choice of δt and Δt, (2) different types of basis functions $U_{\underline{k}}$, (3) automated choice of the number of OPs, and (4) compute $\Psi_{\underline{k}}$ at the beginning and end of the δt MD interval and use them for simple Euler extrapolation, or compute $\Psi_{\underline{k}}$ multiple times during the δt MD interval and then smooth the $\Psi_{\underline{k}}$ profiles via least squares fitting to better estimate $d\Psi_{\underline{k}}/dt$ for carrying out the extrapolation to $t + \Delta t$.

Several cautionary notes are in order. The deterministic forces of the Langevin equations contribute to $\Pi_{\underline{k}}$ (associated momenta of $\Psi_{\underline{k}}$) to order Δt, while the random forces contribute to order $O(\Delta t^{1/2})$. Thus, the cumulative effect of the random Langevin force can be overestimated by the MD/OPX scheme. This difficulty is overcome by proper choice of δt and Δt. Another source of error is that δt might not be large enough. If δt is too short, this would in effect violate the assumed equivalence between the long-time and ensemble averages on which our derivation of the Langevin equations is based[1]. We address this by performing simulations to determine an acceptable δt, for which results become insensitive to further δt increases. Thus δt cannot be too short, but the extrapolation scheme requires that δt also be short relative to the characteristic time for which the OPs change appreciably. Thus, one must make use of the adaptive choice of δt and Δt provided in MD/OPX to achieve an acceptable balance between accuracy and efficiency.

IV Application to CCMV Capsid Structural Transitions (STs)

In this section, the background of CCMV capsid STs is briefly reviewed. Computations of structural OPs generated with Legendre polynomials are shown to capture the nanoscale dynamics of capsomers in a CCMV capsid during its expansion. An exploratory MD/OPX simulation on the swollen form of CCMV capsid for 1ns is analyzed quantitatively and the results are shown to agree with those of a direct MD simulation. The MD/OPX performance is presented along with simulation timing results and its potential optimization is discussed.

IV.1 CCMV Capsid STs

CCMV is a member of the bromovirus group of the Bromoviridae family. Its genome consists of four positive-sense single-stranded RNA molecules, two of which are encapsulated separately in two virions and the remaining two form a third type of particle together. Because the purified RNA and coat protein of CCMV can reassemble in vitro to produce infectious virions, CCMV is an excellent system for studying protein-protein and protein-RNA interactions, which provide important information in the assembly and disassembly of icosahedral viruses, ribosomes and other bionanosystems.

The crystal structure of wild-type CCMV was solved at 3.2Å resolution by X-ray crystallography[2]. Its capsid is comprised of 180 chemically identical protein subunits that form a 286Å diameter icosahedral shell displaying a T=3 quasi-symmetry. Each protein subunit is composed of 190 amino acids taking three quasi-equivalent positions on the capsid surface; one copy of the three quasi-threefold related subunits, called a protomer, is shown in Fig. 2a with the three subunits colored in blue (A), red (B), and green (C). The native capsid (Fig. 2b) is organized in 12 pentameric and 20 hexameric capsomers. There are 5 A subunits in each pentamer and 3 B and 3 C subunits in each hexamer.

Like many plant viruses, the morphology and stability of CCMV is affected by conditions in the host medium (e.g. pH, temperature and ionic strength)[2,4]. Native CCMV is stable in a compact form around pH 5.0. When pH is raised to 7.0 at low ionic strength (< 0.2 M) in the absence of divalent cations, the capsid undergoes a 10% radial expansion at the quasi-threefold axes. In the expansion scheme proposed by Liu et al.[4], the swollen capsid (Fig. 2c) can be generated by taking the pentamers and hexamers through the following rigid-body changes from their native configurations: translate pentamers by 24Å radially and rotate them counter-clockwise by 9° around their 5-fold axes; and translate hexamers by 21Å radially and rotate them counter-clockwise by 8° around their 3-fold axes. This scenario provides a good test for our MD/OPX methodology.

IV.2 OPs capturing capsomer dynamics during CCMV capsid expansion

To demonstrate the viability of our OPs, 4 intermediate structures and the final swollen one were generated from native CCMV capsid equally distributed along its expansion path and OPs generated with orthonormalized Legendre polynomials for the 5 atomic structures were calculated by setting the native capsid as the reference configuration. Results show that our structural OPs are able to capture the nanoscale dynamics of the capsomers, i.e. radial translation and rotation around their symmetric axes, during the capsid expansion as follows.

Starting from the native CCMV capsid, 5 steps were applied to transform the native capsid structure into the swollen structure with 4.8Å-translation and 1.8°-rotation for pentamers and 4.25Å-translation and 1.6°-rotation for hexamers in each step. The inner shell view of the 5 result structures are shown in Fig. 3 (a1) to (a5)). With these 5 atomic structures, we calculated the OPs using Eqn. (II.5) and then the coherent contribution to their atomic coordinates, i.e., the first term on the RHS of Eqn. (II.4). These coherent structures generated with OPs were compared with the 5 original structures to investigate the performance of the OPs.

Four sets of calculations were performed to evaluate the effects of varying the number of OPs, notably 3^3, 4^3, 5^3 and 6^3. For n^3 OPs, Legendre polynomials of order $(0,1,\cdots,n-1)$ in the X, Y and Z directions were used (see Sect. II). Fig. 3 (b1) to (b5) show a sample set of structures visualized with the coherent contribution of atomic coordinates calculated from 3^3 OPs. They are found to reflect the overall expansion of the capsid, although the openings formed around the quasi-threefold axes during the expansion are lost. As the number of OPs is increased from 3^3 to 6^3, the calculated coherent structures become closer to their original atomic structures.

To focus on the translation and rotation of capsomers, one pentamer and one hexamer were extracted from the 5 capsid structures along the expansion pathway, as well as their coherent structures calculated from 3^3, 4^3, 5^3 and 6^3 OPs. The center of mass (CM) positions of the protein subunits included in the pentamer from these structures were visualized in Fig. 4a, and similarly for the hexamer in Fig. 4b. In the figure, the black bead represents the CM position of the capsid, which stays at the origin for all structures; blue beads represent CM positions of the protein subunits in the chosen pentamer and hexamer extracted from the 5 original capsid structures; while the red, green, cyan and purple beads correspond to the protein subunits from the 5 coherent structures calculated from 3^3, 4^3, 5^3 and 6^3 OPs. Top view of the blue beads slightly off the symmetry axes of the pentamer and hexamer in Fig. 4 (a) and (b) shows their radial translation and rotation accompanying the capsid expansion. The trajectories of the other 4 sets of beads show that the translation of the pentamer and hexamer during the capsid expansion can be captured with just 3^3 OPs (red beads) and the accuracy increases with the number of OPs, and partial of the pentamer and hexamer rotation is captured with 5^3 (cyan beads) and 6^3 (purple beads) OPs while it is not revealed with 3^3 (red beads) or 4^3 (green beads) OPs. This can be explained by noting that CCMV capsid has 12 pentamers and 20 hexamers and their total translation and rotation degrees of freedom (DoF) are both 96, which is smaller than 5^3 and 6^3, but greater than 3^3 and 4^3. When the number of OPs is smaller than the total translation DoF of capsomers (i.e., preserved structural units defined here), partial of the capsomer translation is captured and when the number of OPs exceeds the translation DoF, the capsomer rotation starts to be captured. This suggests that our OPs capture the ST dynamics when enough OPs are used. The positions and orientations of beads in a coarse-grained model are in a sense a subset of our OPs. However, as we construct our OPs from the all-atom configuration of the nanosystem, universal interatomic force fields can be used for our MD/OPX simulations without recalibration of force field parameters as needed for coarse-grained models. Furthermore, our approach is free from assumptions on the identity of structural units, i.e. forces and energies dictate which structural units are preserved along the transition pathway. This greatly enhances the predictive potential of our methodology over coarse-grained phenomenological bead models.

IV.3 Comparison of MD and MD/OPX Simulations on CCMV Capsid

Even though the native CCMV capsid in an aqueous medium is known to undergo a swelling process in response to changes in host medium conditions, direct MD simulations show that when the *N*-terminal residues are omitted from the capsid proteins by only using the X-ray structure downloaded from PDB (ID: 1CWP), the native CCMV capsid does not undergo significant structural changes over 10ns in vacuum. To study the evolution of our OPs of Sect. II and demonstrate our MD/OPX methodology, we started with the swollen CCMV capsid structure constructed as described above and then simulated its ensuing shrinkage in vacuum. For the MD/OPX simulation, Legendre polynomials of order (0, 1, 2) were used to generate 3^3 OPs for capturing capsid shrinkage dynamics. 1000 MD/OPX cycles were run with each cycle composed of 100 1fs-MD steps, i.e. δt=100fs and one OPX of 900 equivalent MD steps, i.e. Δt=1000fs. Results, including snapshots of the simulated trajectories, time courses of the capsid radii, root mean square deviation (RMSD) from the starting structure, individual OPs and program timing are compared with those of a direct MD simulation as follows.

The structures after 1ns simulation using MD (a) and MD/OPX (b) are shown in Fig. 5. To quantitatively compare them, the profile of the average, minimum and maximum radii of the capsid backbone (the distance between one of the protein backbone atoms and the capsid CM) over time are presented in Fig. 6. These radius values for MD/OPX and MD simulation are seen to be in good agreement. In both simulations, the average capsid radius is decreased from 141Å to about 136Å (i.e., by 3.5%). The contributions from MD and OPX to the overall capsid shrinkage in the MD/OPX simulation are computed as 85.2% for OPX and the rest 14.8% for MD (values are obtained through averaging over 1000 “MD-OPX” cycles). Given the fact that each MD/OPX cycle is composed of 100 1fs-MD steps and one OPX of 900 equivalent MD steps, the capsid shrinkage captured by OPX (85.2%) is close to its simulated time portion (90%) and the contribution is significant. Note that the contribution from OPX to the bionanosystem dynamics in MD/OPX simulations is proportional to its simulated time portion, i.e., as the ratio of the OP timestep to the MD run time increases in each cycle, the contribution from OPX becomes more significant. Fig. 7 shows time courses of the RMSD of atomic positions between the simulated CCMV capsid structure along the trajectory and its starting structure using MD and MD/OPX. There is only a small difference between the two curves. Both show that the RMSD is increased to about 6Å in 1ns and there is still an increasing trend at the end of the simulation, indicating the capsid will continue undergoing significant changes after 1ns.

With the above comparison, predictions of the MD/OPX simulation for shrinkage of the swollen CCMV capsid in vacuum are found to agree well with those of MD. Finally, even though only 3^3 OPs were used, the agreement was excellent. This shows that while the OPs may not capture all the details, the residuals (see Eqn. II.4) retain the details lost by the OPs. This suggests that completeness of the set of OPs is not a concern as long as the residuals needed to correct the coherent dynamics are not too large. In the results cited above the residuals satisfy this criterion (see below).

IV.4 Characteristics of OPs and MD/OPX

As shown above, the extrapolation of OPs works well to evolve the nanostructure in time. To validate such OPX, the 3^3 OPs generated with Legendre polynomials of order (0, 1, 2) in X, Y, and Z directions for the 1ns MD simulation output trajectory of the swollen CCMV capsid are plotted versus time. Fig. 8a shows time courses of the capsid OPs labeled with indices (1, 0, 0) in the X-direction, (0, 1, 0) for Y and (0, 0, 1) for Z (X100, Y010 and Z001), reflecting an isotropic shrinkage of the capsid. The OPs vary slowly and can be readily extrapolated to long time on a timescale of ns. Other OPs shown in Fig. 8b are seen to fluctuate rapidly around zero over 1ns. A closer look at these OPs from 60ps to 100ps (Fig. 8c) shows that their values are comparatively stable on a timescale of 10ps, which implies their characteristic time periods are 10ps or longer. This means that all the OPs display much less stochastic behavior than do individual atoms and have much longer characteristic time than the 10^{-14} second timescale of fast atomic vibrations and collisions as expected; thus the OPs satisfy one of the criteria for the applicability of MD/OPX. The origin of the OP fluctuations in Fig. 8b is likely that they indicate the elastic vibrations of the capsid in free space. We expect that when the capsid is placed in an aqueous medium, frictional effects will dampen these fluctuations so that the higher order OPs have longer characteristic time and can also be extrapolated over larger time intervals.

Another basis for MD/OPX is identified by investigating the coherent and residual contributions to the atomic configuration along its simulated trajectory. Fig. 9 shows a snapshot of the CCMV capsid after 1ps MD/OPX simulation with atomic coordinates given by their coherent coordinates plus the residuals. Similar to the results presented in Part IV.2, the coherent coordinates calculated with 3^3 OPs are found to reflect overall capsid structure with the residuals close to zero. This suggests that the slow collection motions in viral STs are captured with coherent coordinates computed through the slowly-varying OPs while the fast motions including atomic vibrations and collisions are accounted for by the rapidly-fluctuating residuals.

With the characteristic time of OPs determined to be in the range of 10ps to 1ns, the adaptive procedure described in Sect. III has been implemented to find the timestep of OPs allowed for MD/OPX simulations. While it varies with other simulation parameters, for a simulation using 100 1fs-MD steps (i.e., δt_0=100fs) to anneal the result structure of OPX and keeping δt as 100fs, the OP timestep Δt displays a typical value of 1.8ps. With another additional 1000 energy minimization steps after OPX to anneal the structure, the optimized OP timestep becomes 5ps. Further optimization of all run parameters is under investigation.

To test the stability of MD/OPX, a 16ns simulation was run on the swollen CCMV capsid and the results (Fig. 10) show that the average radius of the capsid backbone decreases from 141Å to 132Å, the minimum capsid radius drops comparatively fast in the beginning of the simulation from 116Å to 110Å in 2ns and decreases more slowly for the rest of the time to 103Å at 16ns, while the maximum capsid radius keeps fluctuating with an overall decreasing trend from 166Å to 163Å. As a consequence, the capsid thickens from 50Å to 60Å during its shrinkage. The change in the rate of capsid shrinkage can be explained by the fact that with the swollen capsid approaching its native state near equilibrium, the rate of progression decreases. This also proves that our MD/OPX approach is stable for long-time bionanosystem simulations.

IV.5 MD/OPX performance and potential optimization

The simulations in this study were run on the Big Red cluster at Indiana University with 64 IBM processors for parallel MD and 1 processor for the serial OPX code. Timing results show that about 0.097 seconds are needed for one MD step and 15 seconds for one OPX step. As a result, the OPX in each cycle is 372 times faster than the MD run for the 900 equivalent MD steps. For a 1ns simulation, MD/OPX is 9.76 times faster than the direct MD.

Because the overall speedup of MD/OPX over MD is directly proportional to the $\Delta t/\delta t$ ratio, the above performance results can be improved by a factor of 5 by increasing the OP timestep to 5ps as obtained from the adaptive MD/OPX simulations. However, additional energy minimization and short MD run are needed in the present implementation to anneal the OPX result structure and they downgrades the program running performance. The present implementation of MD/OPX makes use of the Tcl scripting in NAMD. Atomic structures are transferred inefficiently via file saving and reading between NAMD and the Fortran OPX program. And for the adaptive procedure, forces on all atoms are loaded to calculate their accelerations as indicators redundantly. These can be avoided by integrating MD/OPX into the core code of NAMD so that the indicators used to judge the structure of

OPX can be queried efficiently and less computation is expected for annealing the structure, i.e., fewer energy minimization steps and shorter MD run. With this and the $\Delta t/\delta t$ ratio obtained as 50, the overall acceleration of MD/OPX over MD is expected reach 50. Also as the OPs we constructed display different timescales (10ps to 1ns), multiple timesteps can be implemented to further optimize the procedure and a larger average timestep for OPs can be obtained to accelerate the simulations.

V Conclusions

A computational approach, MD/OPX, is developed to simulate large bionanosystems. It makes use of a short MD run to estimate the rate of change of OPs constructed automatically with orthogonal polynomials, which is then used to extrapolate the state of the system over long time. The approach is based on our AMA method for dynamical nanosystems. The all-atom formulation enables the development of a universal simulator, avoiding the need for recalibration with each new application. Thereby an understanding of bionanosystems is achieved based on the principles of molecular physics. The approach greatly extends the realm of applicability of standard MD packages. Unique features of MD/OPX as implemented include:

- automated construction of a set of OPs capturing the structure of complex bionanosystems and their nanoscale dynamics;
- a rigorous demonstration of the timescale separation between the characteristic time of the OPs and atomistic dynamics, and the implied existence of Langevin equations for their stochastic dynamics;
- capturing all-atom detail in whole-nanostructure simulations; and
- optimization of a dual timestep evolution algorithm via dynamic adaptation of simulation run parameters.

The OPs constructed with orthonormalized Legendre polynomials were demonstrated to be capable of capturing the nanoscale dynamics of capsomers during native CCMV capsid expansion. MD/OPX is then tested on shrinkage of the swollen CCMV capsid in vacuum. The simulation results are shown to agree well with those of a direct MD simulation. As implemented here, MD/OPX is 9.76 times faster than NAMD. With potential optimization, the overall acceleration of MD/OPX over NAMD can be improved to about 50 for simulating the CCMV capsid in vacuum. The slowly varying characteristics of OPs revealed and verified in this study justify the development of a direct solver of the Langevin equations for the OPs. A theoretical speed-up of the solver over MD by a factor of $O(N)$ is expected for large bionanosystems composed of N atoms with larger timesteps applied for the time integration and a much smaller number of unknown variables to solve, though this theoretical speed-up is overestimated due to the computations needed for constructing the frictions coefficients and thermal average forces in the early rapidly changing stage of the OP evolution. While the CCMV capsid used to demonstrate our MD/OPX approach has 432,120 atoms, many systems of biological interest are supra-million atoms in size (e.g. large viruses and cell membranes in host media) and as the size of the simulated system increases, the simulation speed-up of our multiscale approach over direct MD will become greater. MD/OPX and the direct solver of the Langevin equations for the OPs based on our AMA methodology hold great promise for long-time simulations of large bionanosystems. Their benefits for health sciences and biotechnology include the computer-aided design of antiviral drugs and vaccines, functionalizing nanoparticles for medical imaging and thermal cancer treatments, and designing nanocapsules (e.g. viral capsids and liposomes) for delivery of therapeutic agents.

Acknowledgments

Thanks to the U.S. Air Force Research laboratories at Wright Patterson Air Force Base for their support through the Student Research Participant Program from the Oak Ridge Institute of Science and Education (ORISE), a seed grant from National Institutes of Health (NIH) through Standford University via SimBios and the support from Indiana University College of Arts and Sciences and the METACyt project through the Center for Cell and Virus Theory.

References

1. Miao Y, Ortoleva P. Journal of Chemical Physics. 2006; 125(4):44901. [PubMed: 16942186]
2. Speir JA, Munshi S, Wang GJ, Baker TS, Johnson JE. Structure. 1995; 3(1):63–78. [PubMed: 7743132]
3. Johnson JE, Speir JA. Journal of Molecular Biology. 1997; 269(5):665–675. [PubMed: 9223631]
4. Liu HJ, Qu CX, Johnson JE, Case DA. Journal of Structural Biology. 2003; 142(3):356–363. [PubMed: 12781662]
5. Liepold LO, Revis J, Allen M, Oltrogge L, Young M, Douglas T. Phys Biol. 2005; (2):S166–S172. [PubMed: 16280622]
6. Speir JA, Bothner B, Qu C, Willits DA, Young MJ, Johnson JE. Journal of Virology. 2006; 80(7): 3582–3981. [PubMed: 16537626]
7. Canady MA, Tihova M, Hanzlik TN, Johnson JE, Yeager M. Journal of Molecular Biology. 2000; (299):573–584. [PubMed: 10835268]
8. Canady MA, Tsuruta H, Johnson JE. Journal of Molecular Biology. 2001; (311):803–814. [PubMed: 11518532]
9. Taylor DJ, Krishna NK, Canady MA, Schneemann A, Johnson JE. Journal of Virology. 2002; 76(19):9972–9980. [PubMed: 12208973]
10. Taylor DJ, Wang Q, Bothner B, Natarajan P, Finn MG, Johnson JE. Chem Commun. 2003:2770–2771.
11. Lee KK, Tsuruta H, Hendrix RW, Duda RL, Johnson JE. Journal of Molecular Biology. 2005; (352):723–735. [PubMed: 16095623]
12. Wikoff WR, Conway JF, Tang J, Lee KK, Gan L, Cheng N, Duda RL, Hendrix RW, Steven AC, Johnson JE. Journal of Structural Biology. 2006:300–306. [PubMed: 16427314]
13. Fricks CE, Hogle JM. Journal of Virology. 1990; 64(5):1934–1945. [PubMed: 2157861]
14. Belnap DM, Filman DJ, Trus BL, Cheng NQ, Booy FP, Conway JF, Curry S, Hiremath CN, Tsang SK, Steven AC, Hogle JM. Journal of Virology. 2000; 74(3):1342–1354. [PubMed: 10627545]
15. Belnap DM, McDermott BM, Filman DJ, Cheng N, Trus BL, Zuccola HJ, Racaniello VR, Hogle JM, Steven AC. Proc Natl Acad Sci. 2000; 97:73–78. [PubMed: 10618373]
16. Tsang SK, McDermott BM, Racaniello VR, Hogle JM. Journal of Virology. 2001; 75(11):4984–4989. [PubMed: 11333877]
17. Hogle JM. Annual Review of Microbiology. 2002; (56):677–702.
18. Arkhipov A, Freddolino PL, Schulten K. Structure. 2006; (14):1767–1777. [PubMed: 17161367]
19. Harries D, May S, Gelbart WM, Ben-Shaul A. Biophysical Journal. 1998; 75:159. [PubMed: 9649376]
20. Speelman B, Brooks BR, Post CB. Biophysical Journal. 2001; 80:121–129. [PubMed: 11159387]
21. Freddolino PL, Arkhipov AS, Larson SB, McPherson A, Schulten K. Structure. 2006; 14(3):437–449. [PubMed: 16531228]
22. Durup J. J Phys Chem. 1991; 95:1817–1829.
23. Tuckerman ME, Berne BJ. Journal of Chemical Physics. 1991; 95(11):8362–8364.
24. Tuckerman ME, Berne BJ, Martyna GJ. Journal of Chemical Physics. 1991; 94(10):6811–6815.
25. Askar A, Space B, Rabitz H. Journal of Physical Chemistry. 1995; 99(19):7330–7338.
26. Reich S. Physica D. 1995; 89:28–42.
27. Phelps DK, Post CB. Journal of Molecular Biology. 1995; 254(4):544–551. [PubMed: 7500332]

28. Phelps DK, Rossky PJ, Post CB. Journal of Molecular Biology. 1998; 276(2):331–337. [PubMed: 9512706]
29. Elezgaray J, Sanejouand YH. Biopolymers. 1998; 46(7):493–501. [PubMed: 9838873]
30. Elezgaray J, Sanejouand YH. Journal of Computational Chemistry. 2000; 21(14):1274–1282.
31. Feenstra KA, Hess B, Berendsen HJC. Journal of Computational Chemistry. 1999; 20(8):786–798.
32. Sorensen MR, Voter AF. J Chem Phys. 2000; 112(21):9599–9606.
33. Chun HM, Padilla CE, Chin DN, Watanabe M, Karlov VI, Alper HE, Soosaar K, Blair KB, Becker OM, Caves LSD, Nagle R, Haney DN, Farmer BL. Journal of Computational Chemistry. 2000; 21(3):159–184.
34. Tama F, Brooks CL. Journal of Molecular Biology. 2002; 318(3):733–747. [PubMed: 12054819]
35. Tama F, Brooks CL. Journal of Molecular Biology. 2005; 345(2):299–314. [PubMed: 15571723]
36. van Vlijmen HWT, Karplus M. Journal of Chemical Physics. 2001; 115(2):691–698.
37. van Vlijmen HWT, Karplus M. Journal of Molecular Biology. 2005; 350(3):528–542. [PubMed: 15922356]
38. Miao Y, Ortoleva PJ. Journal of Chemical Physics. 2006; 125(21):214901. [PubMed: 17166043]
39. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. Journal of Computational Chemistry. 2005; (26):1781–1802. [PubMed: 16222654]
40. Kevrekidis IG, Gear CW, Hummer G. Aiche Journal. 2004; 50(7):1346–1355.
41. Jaqaman K, Ortoleva PJ. J Comput Chem. 2002; 23(4):484–491. [PubMed: 11908085]
42. Shreif Z, Ortoleva P. J Stat Phys. 2008; 130:669–685.

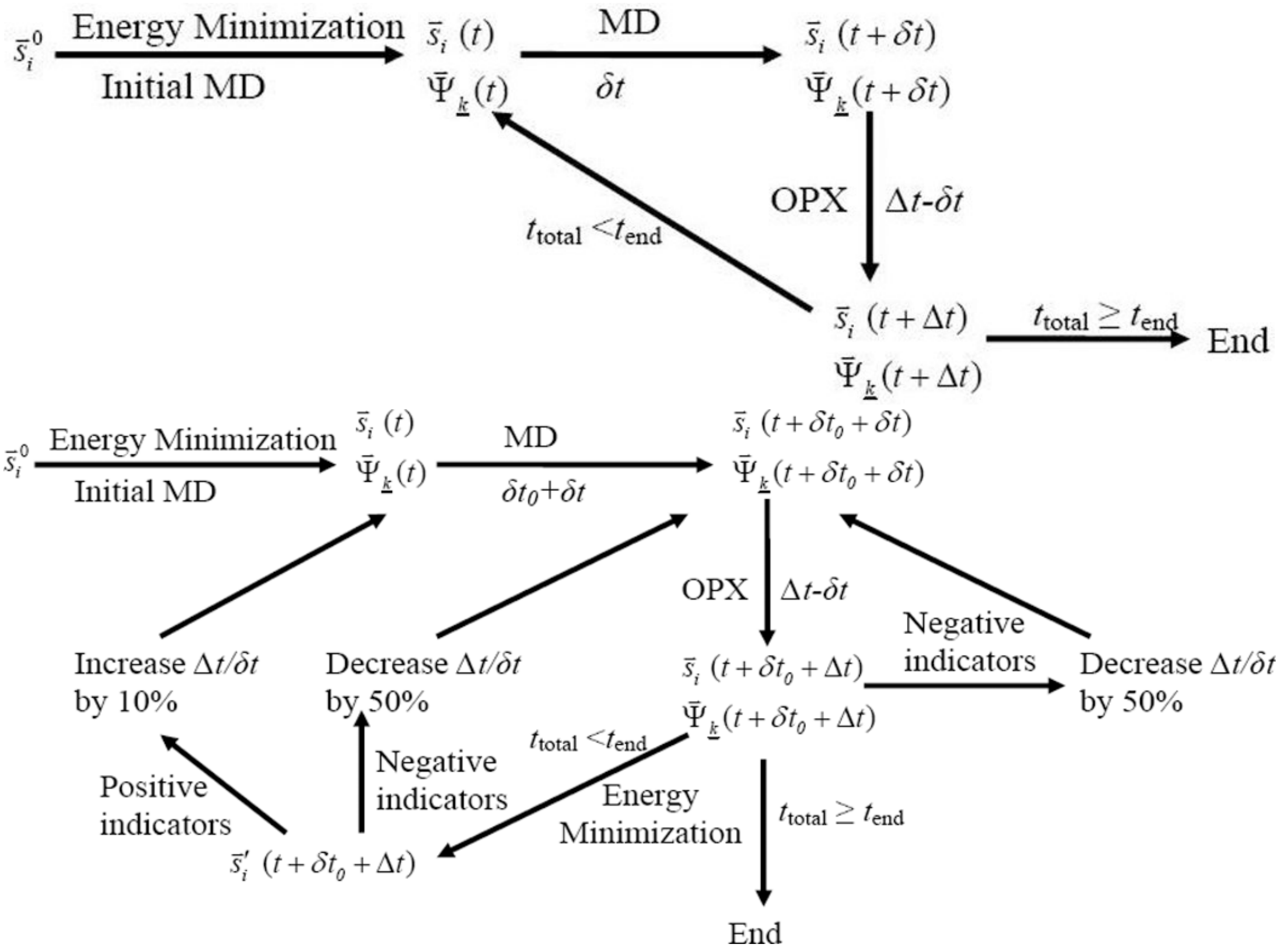
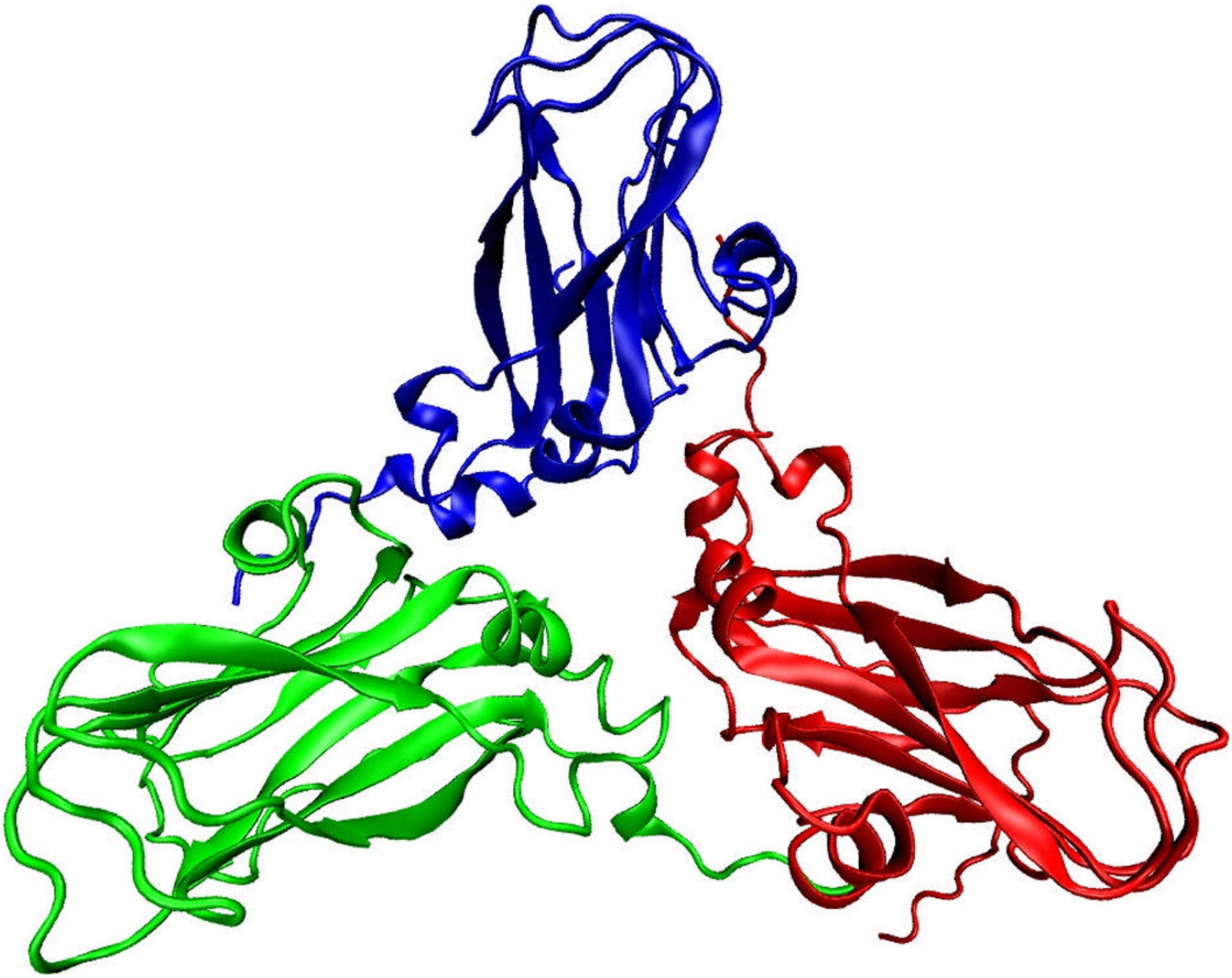
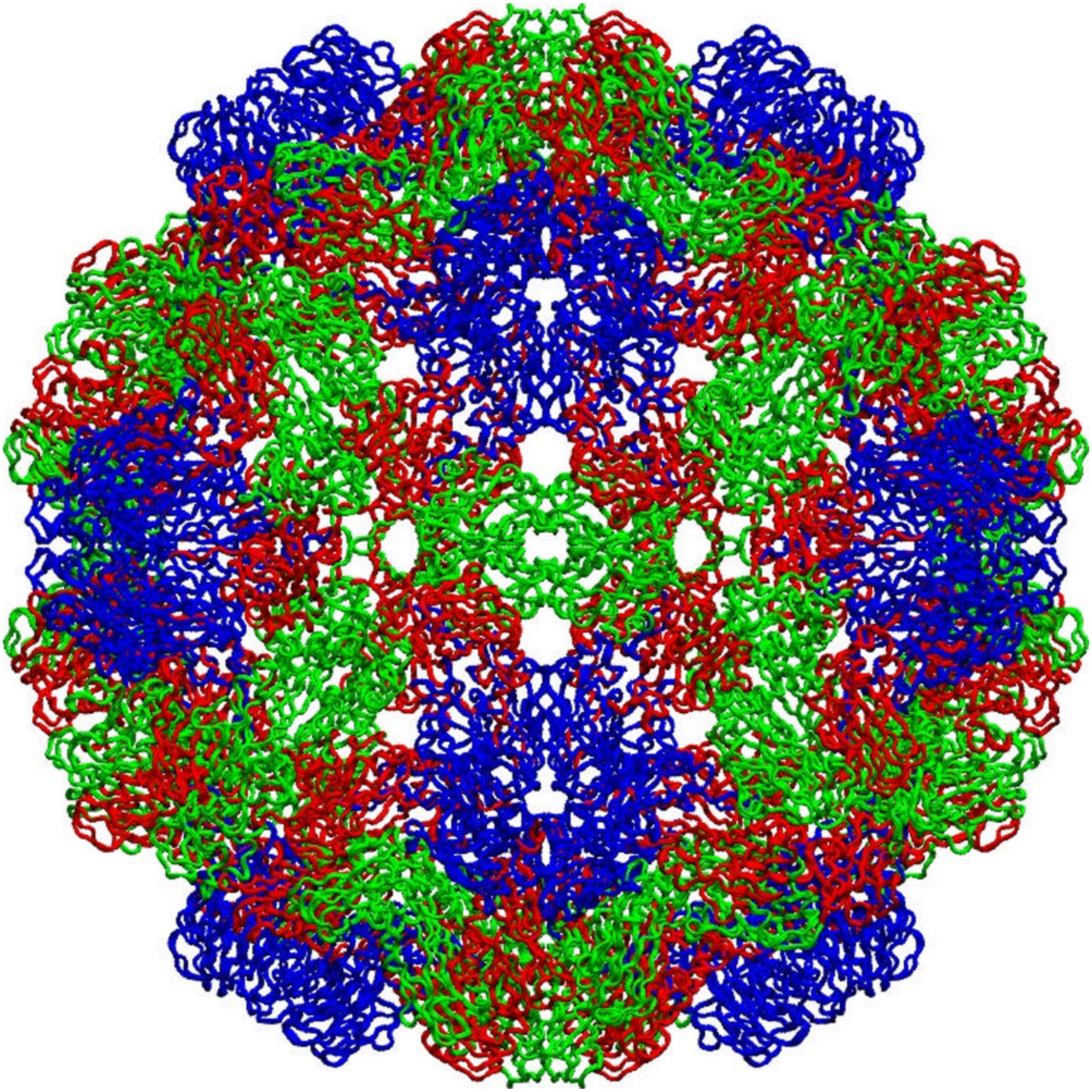

$\bar{s}_i^0$
Energy Minimization
Initial MD
$\bar{s}_i(t)$
$\bar{\Psi}_{\underline{k}}(t)$
MD
δt
$\bar{s}_i(t+\delta t)$
$\bar{\Psi}_{\underline{k}}(t+\delta t)$
OPX
Δt-δt
$t_{\text{total}} < t_{\text{end}}$
$\bar{s}_i(t+\Delta t)$
$\bar{\Psi}_{\underline{k}}(t+\Delta t)$
$t_{\text{total}} \geq t_{\text{end}}$
End
$\bar{s}_i^0$
Energy Minimization
Initial MD
$\bar{s}_i(t)$
$\bar{\Psi}_{\underline{k}}(t)$
MD
$\delta t_0+\delta t$
$\bar{s}_i(t+\delta t_0+\delta t)$
$\bar{\Psi}_{\underline{k}}(t+\delta t_0+\delta t)$
OPX
Δt-δt
Increase $\Delta t/\delta t$ by 10%
Decrease $\Delta t/\delta t$ by 50%
$\bar{s}_i(t+\delta t_0+\Delta t)$
$\bar{\Psi}_{\underline{k}}(t+\delta t_0+\Delta t)$
Negative indicators
Decrease $\Delta t/\delta t$ by 50%
Positive indicators
Negative indicators
$t_{\text{total}} < t_{\text{end}}$
Energy Minimization
$t_{\text{total}} \geq t_{\text{end}}$
$\bar{s}_i'(t+\delta t_0+\Delta t)$
End

Fig. 1.





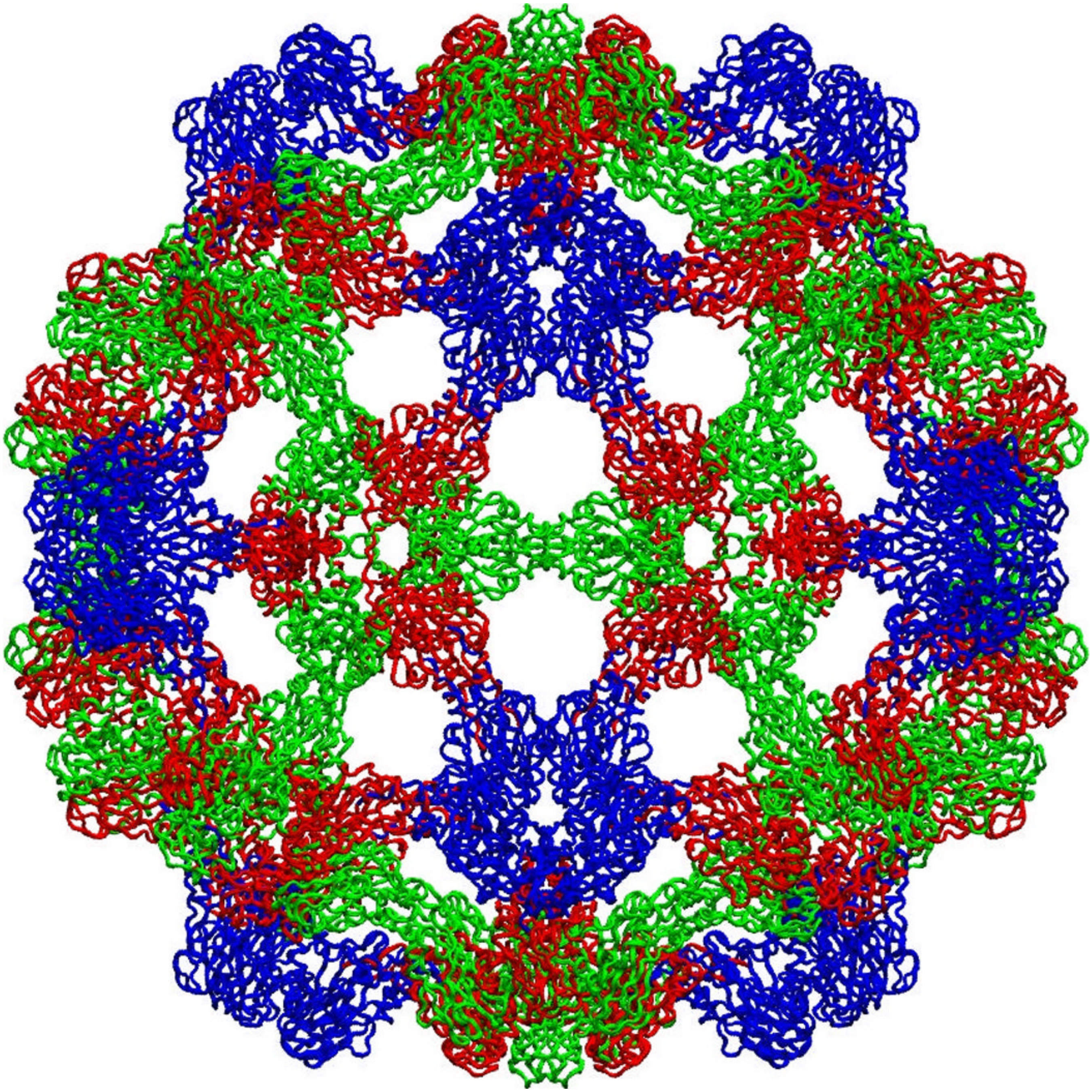
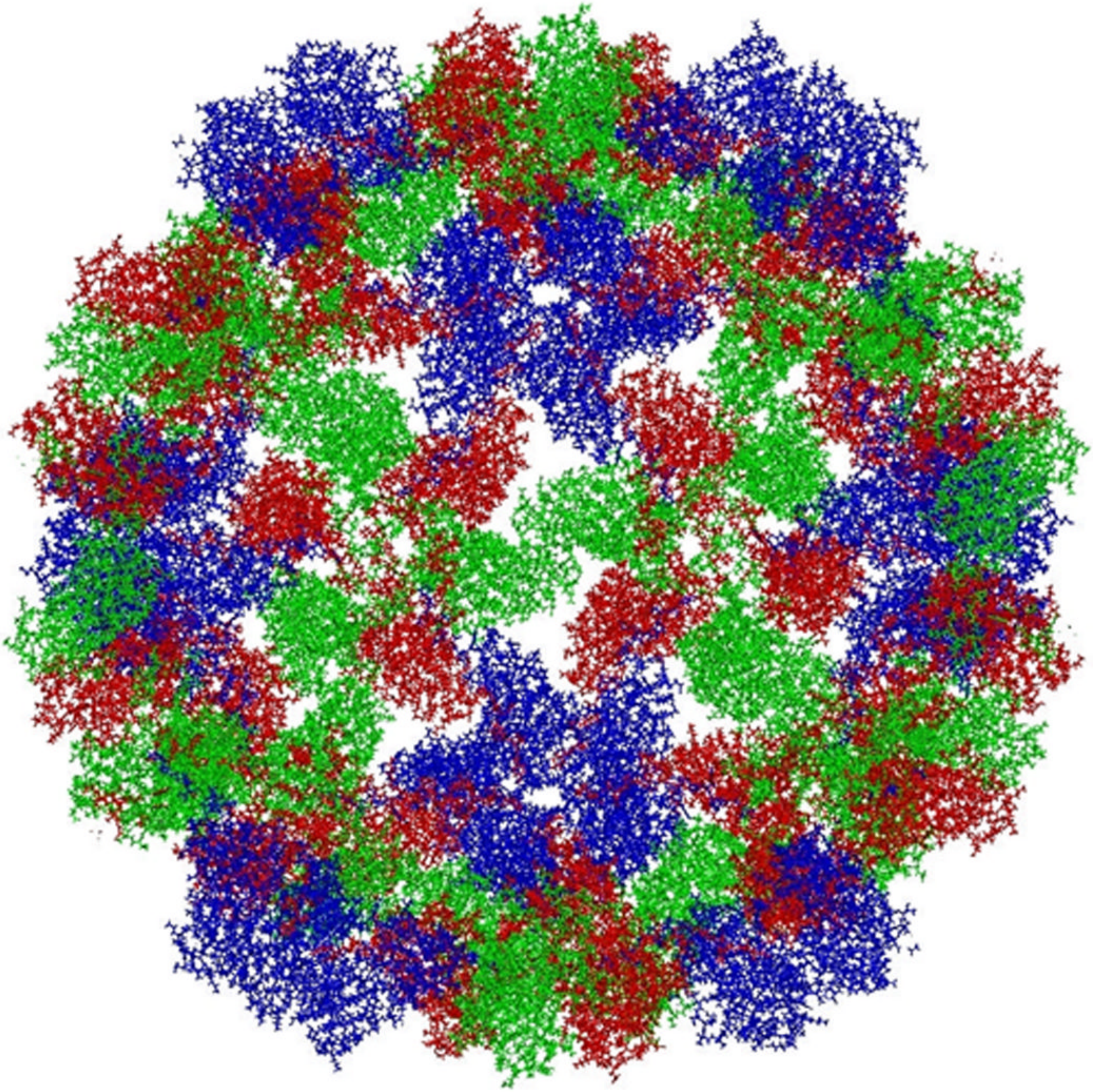
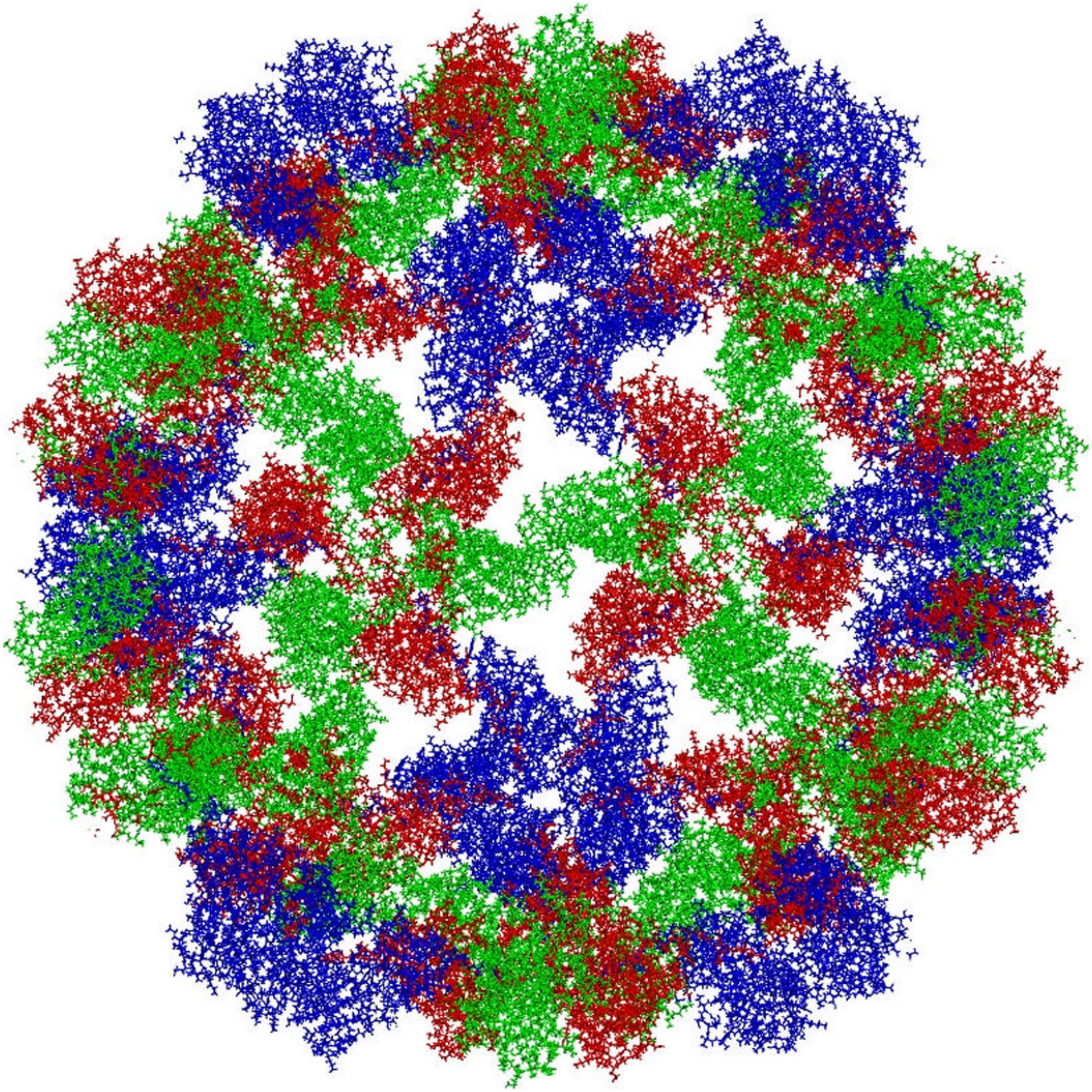
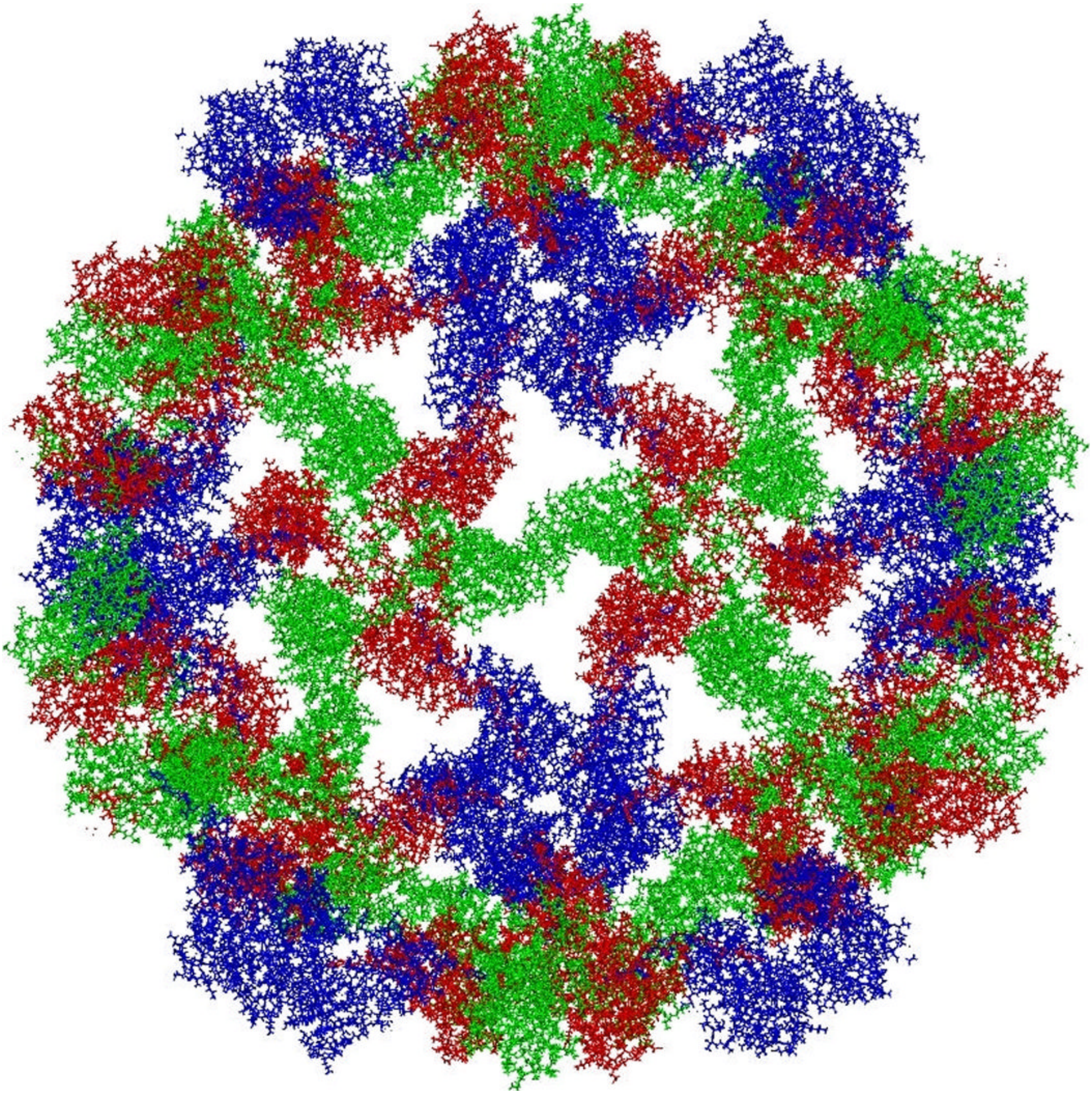
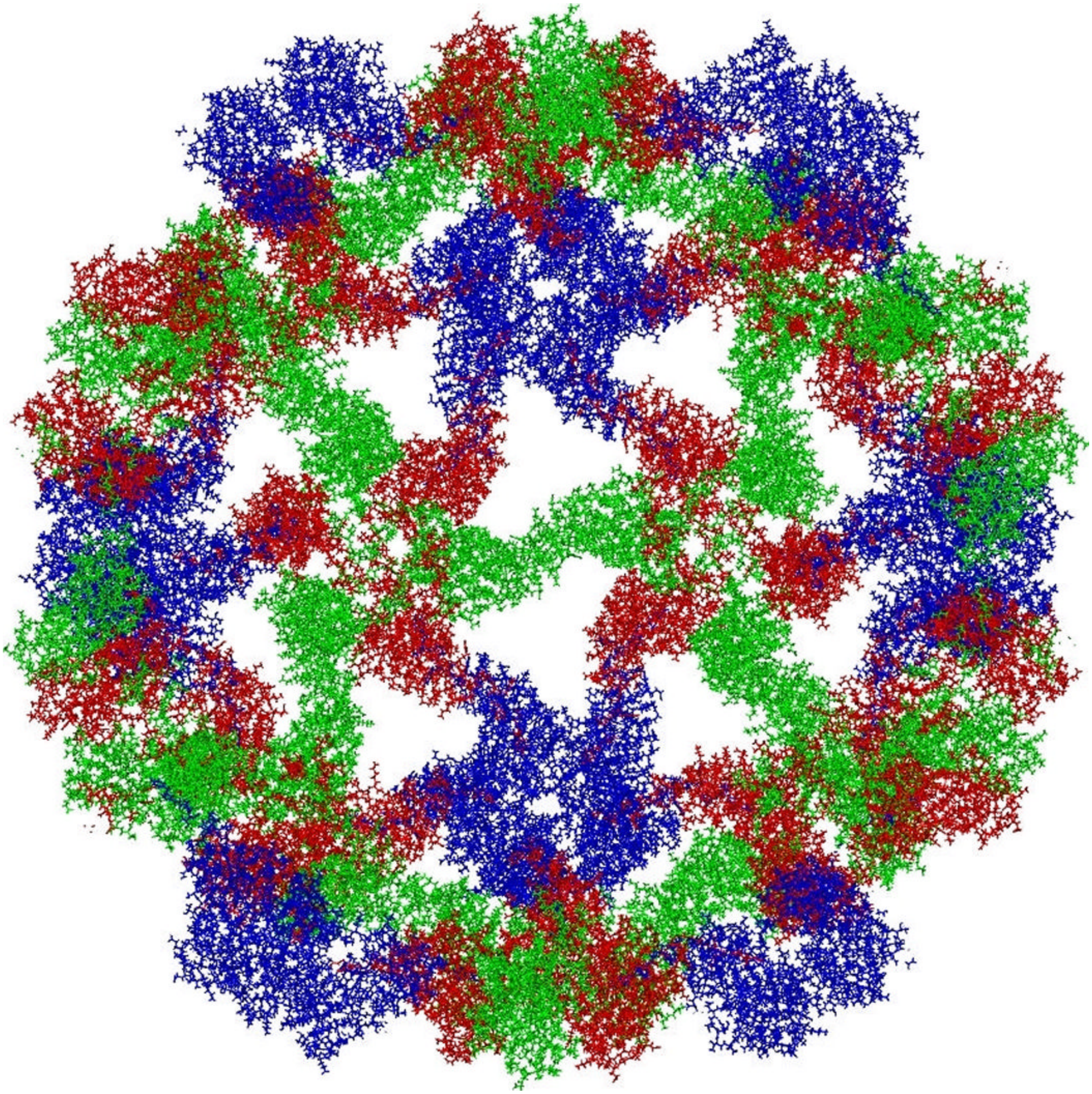


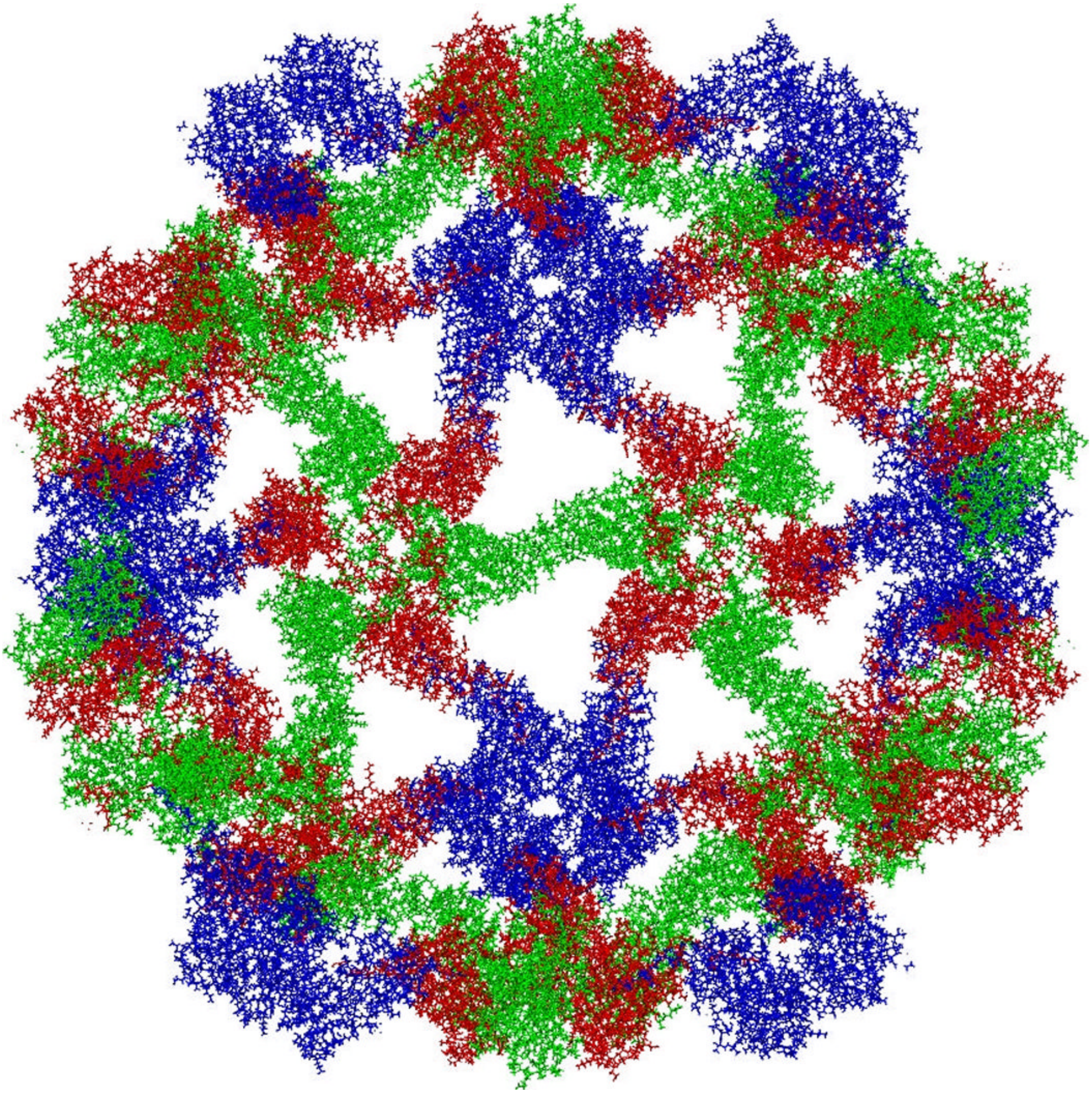
Fig. 2.

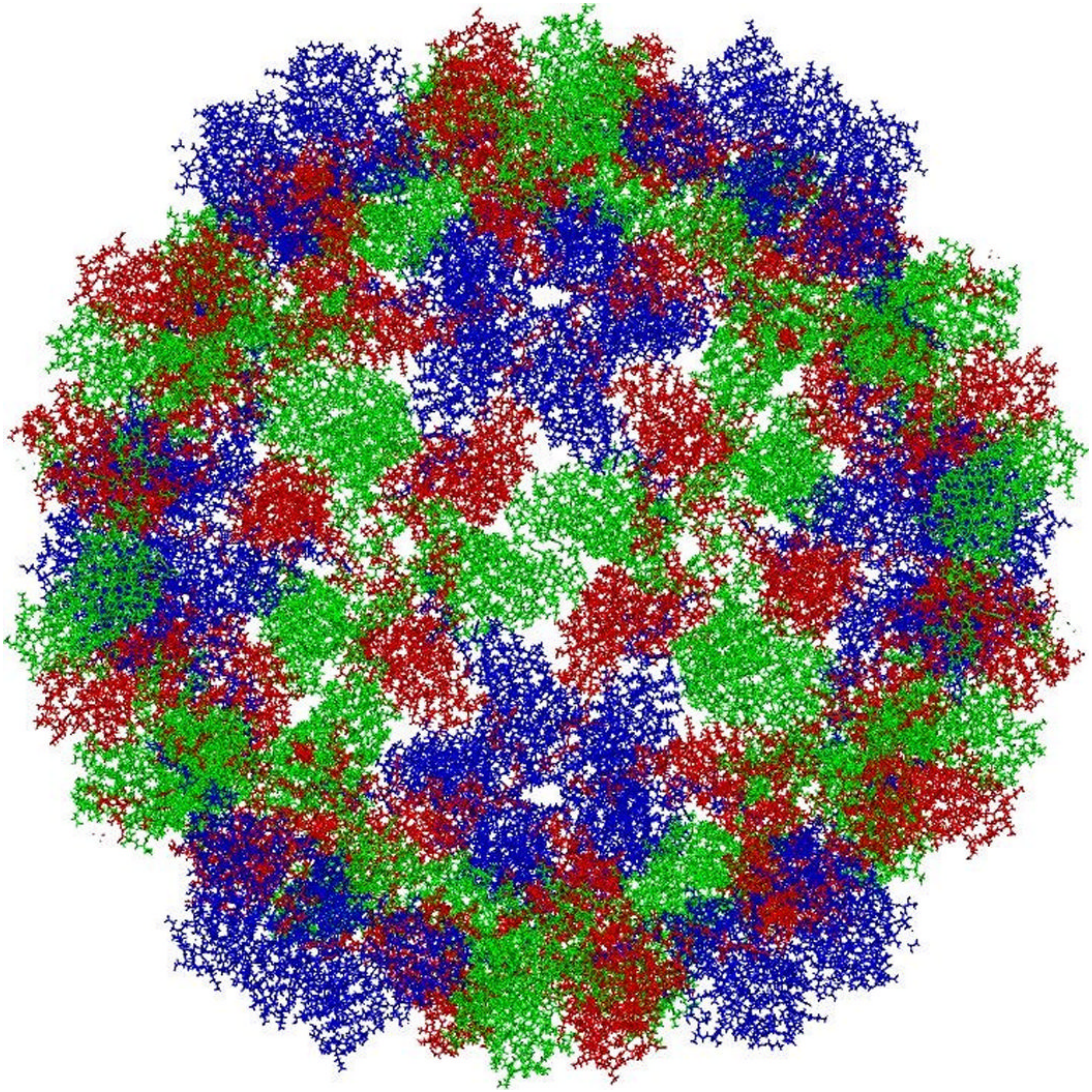


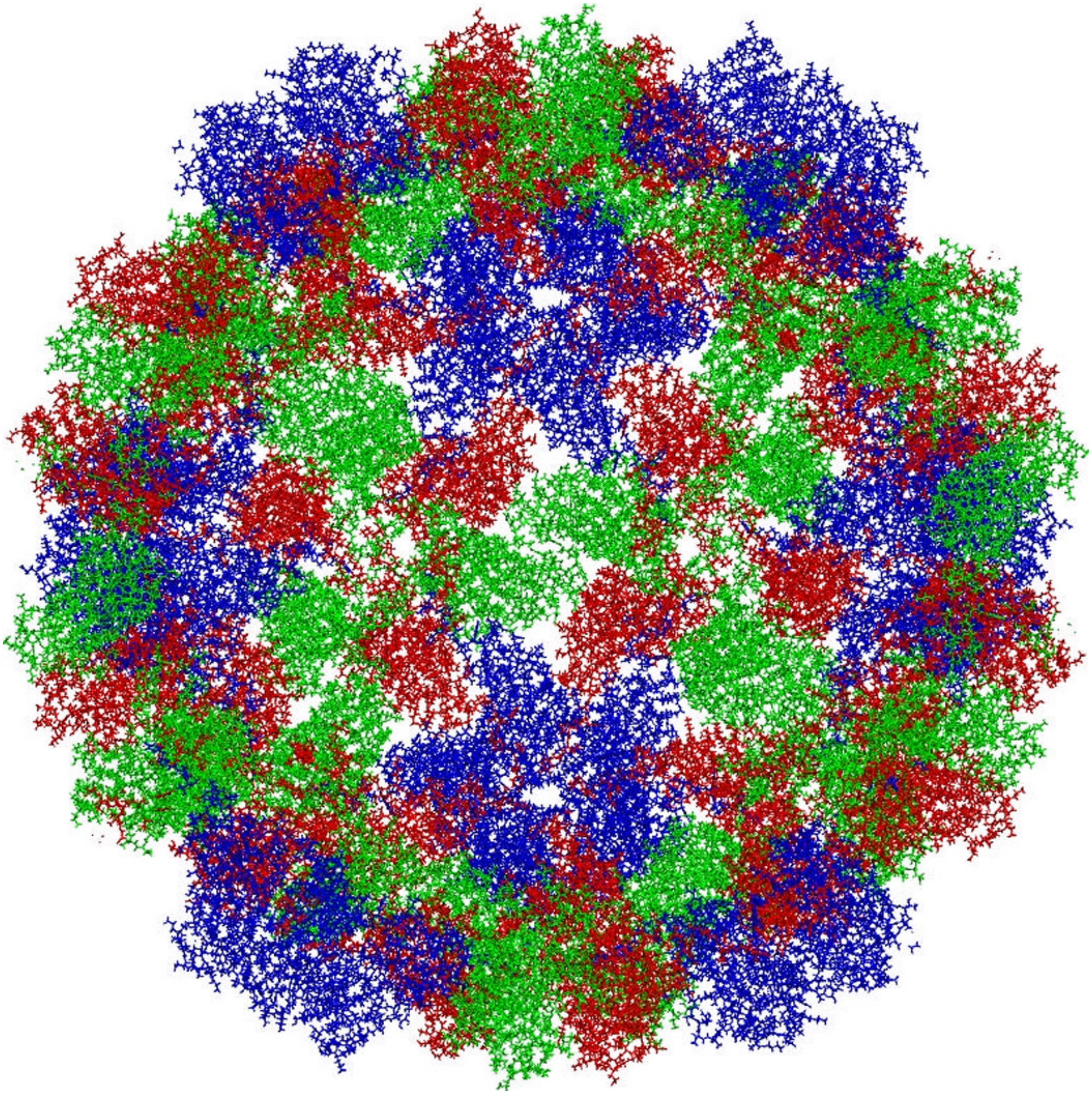


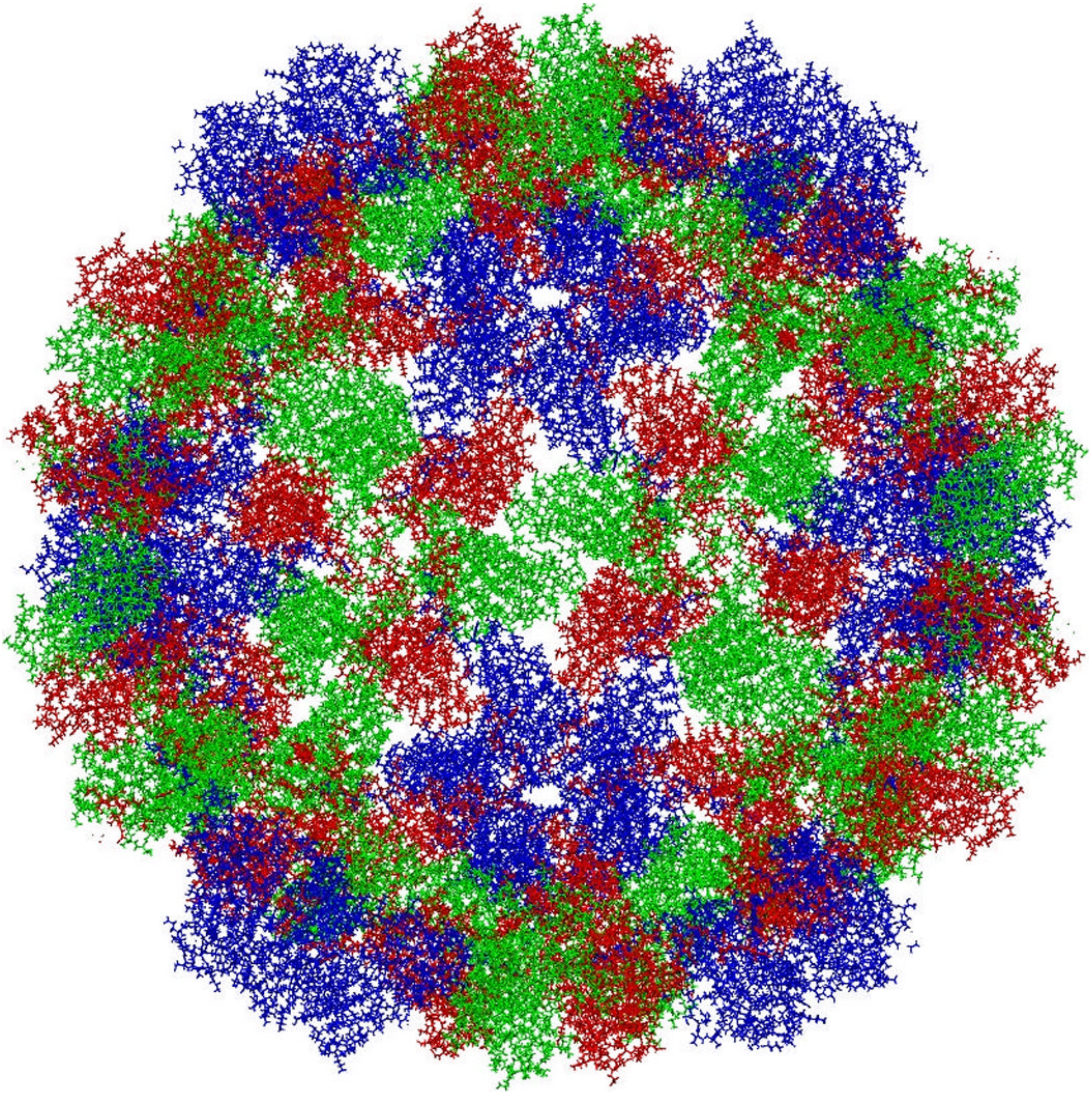


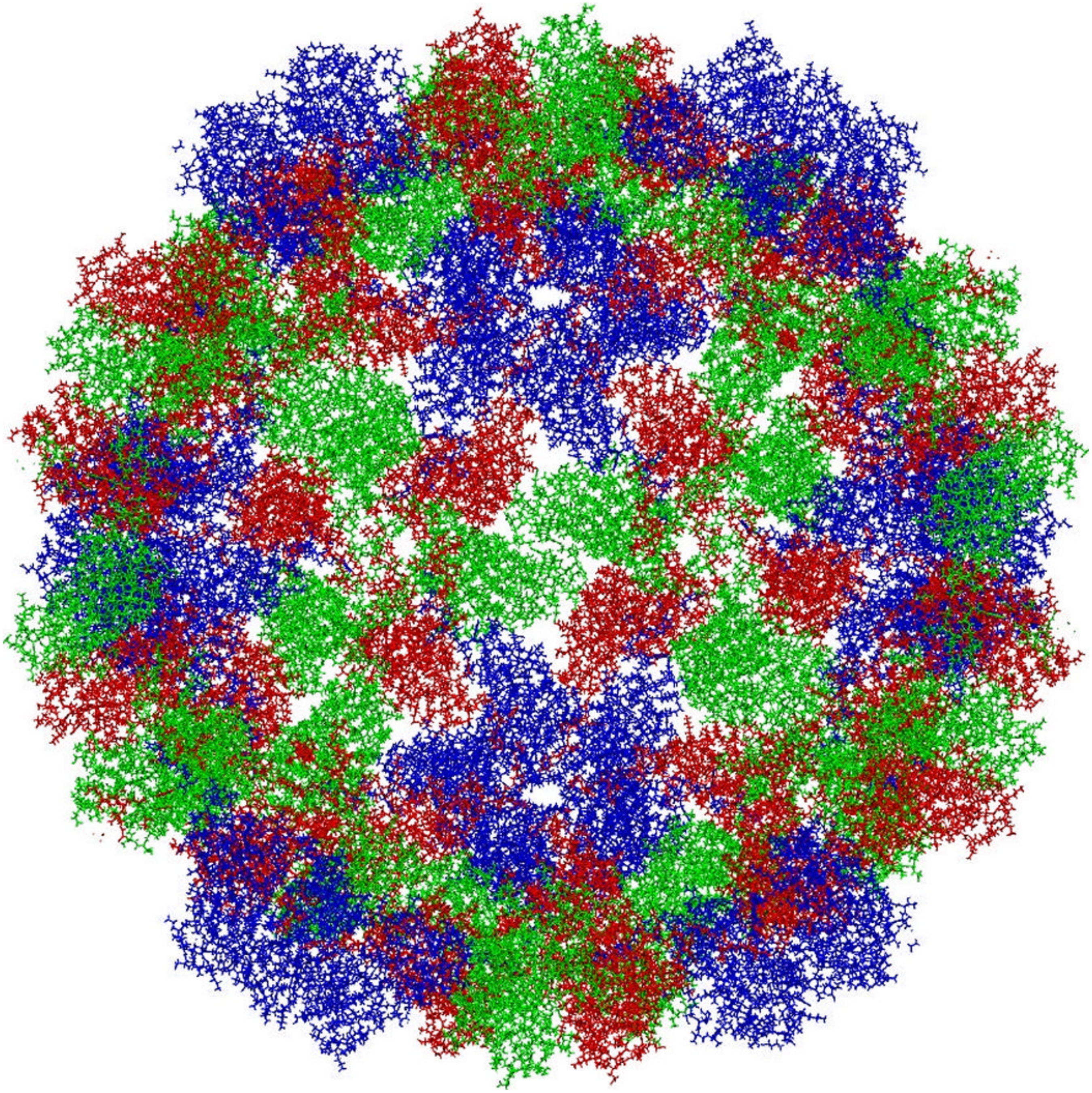












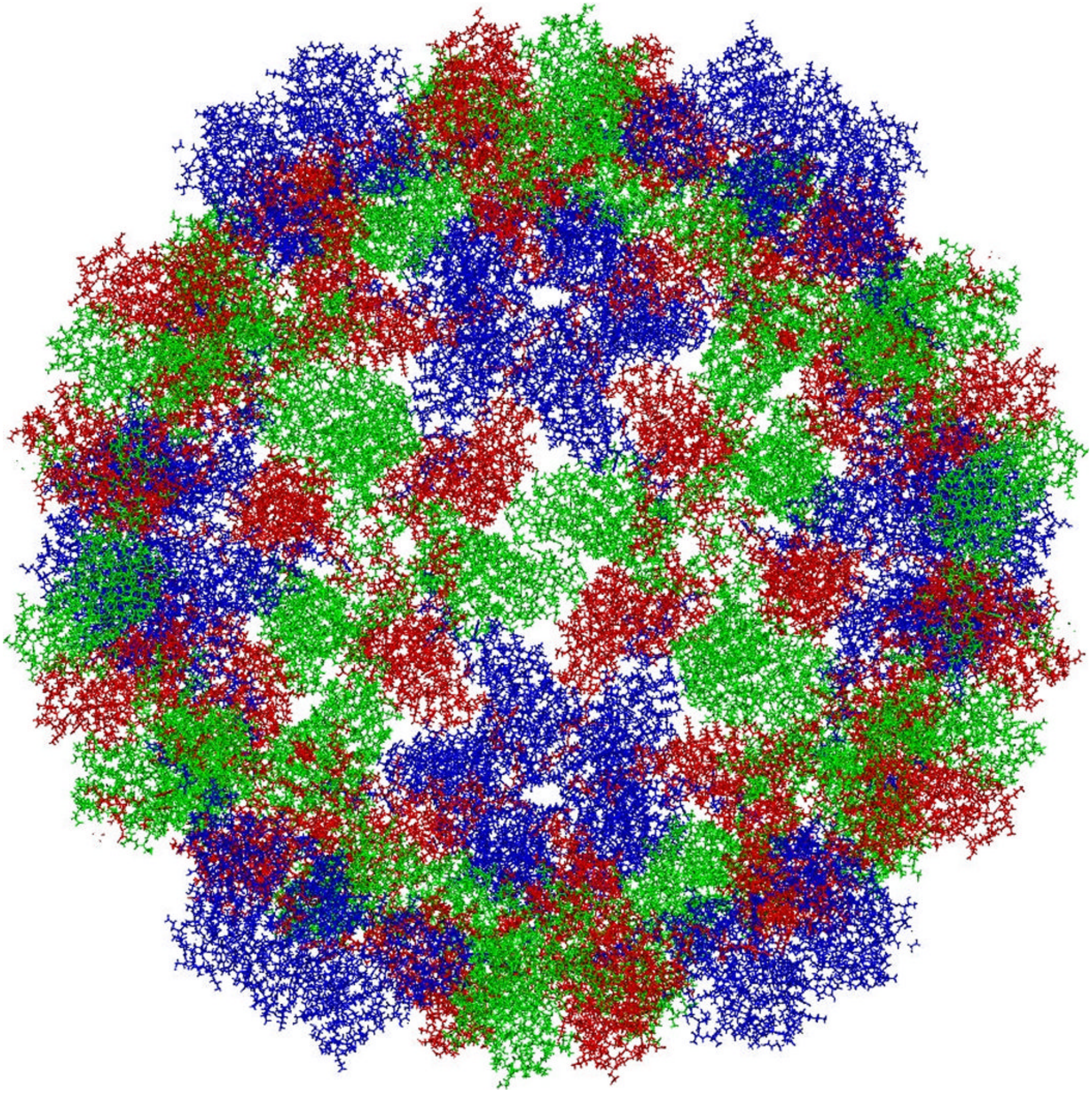
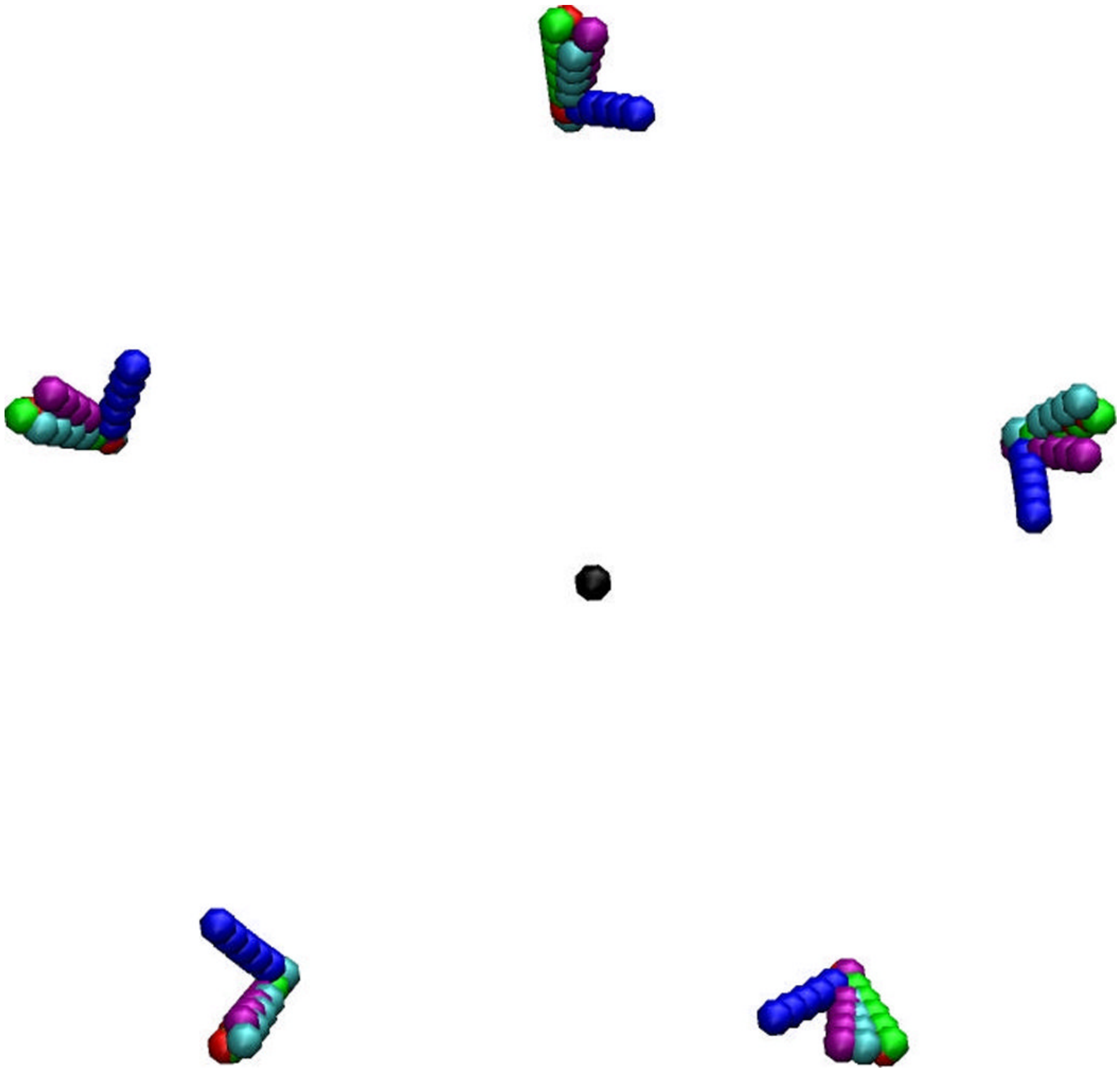


Fig. 3.



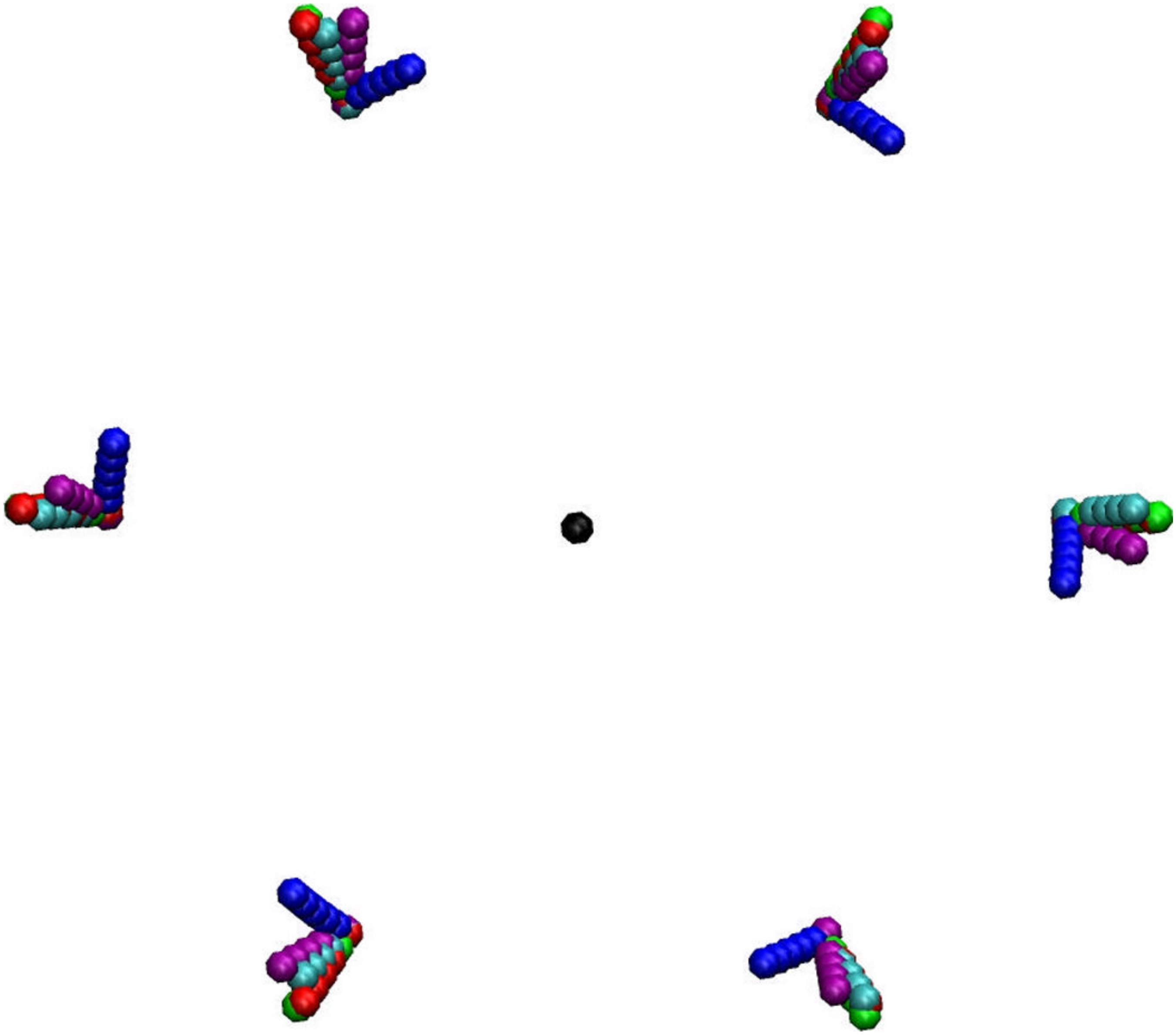
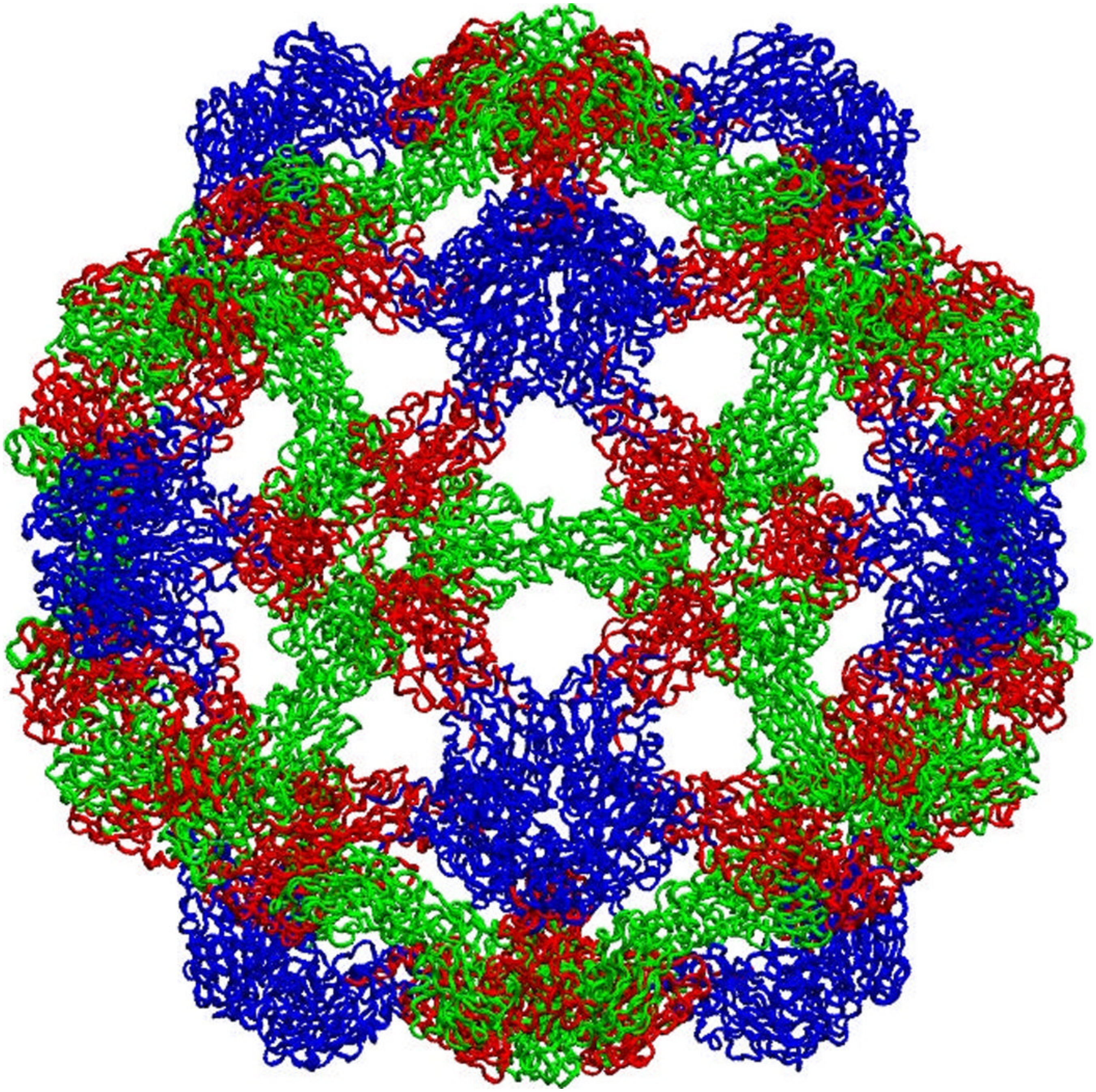


Fig. 4.



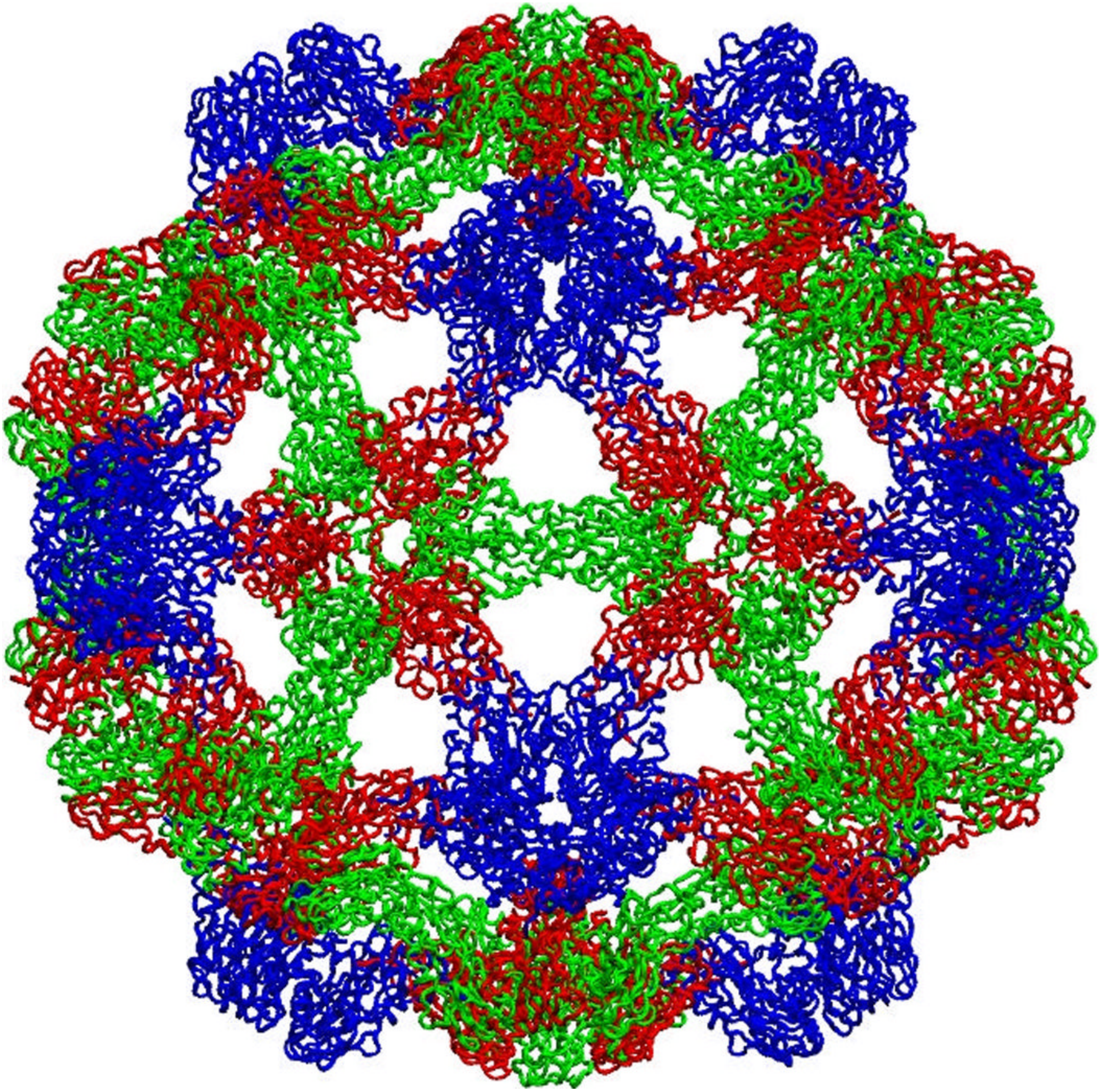


Fig. 5.

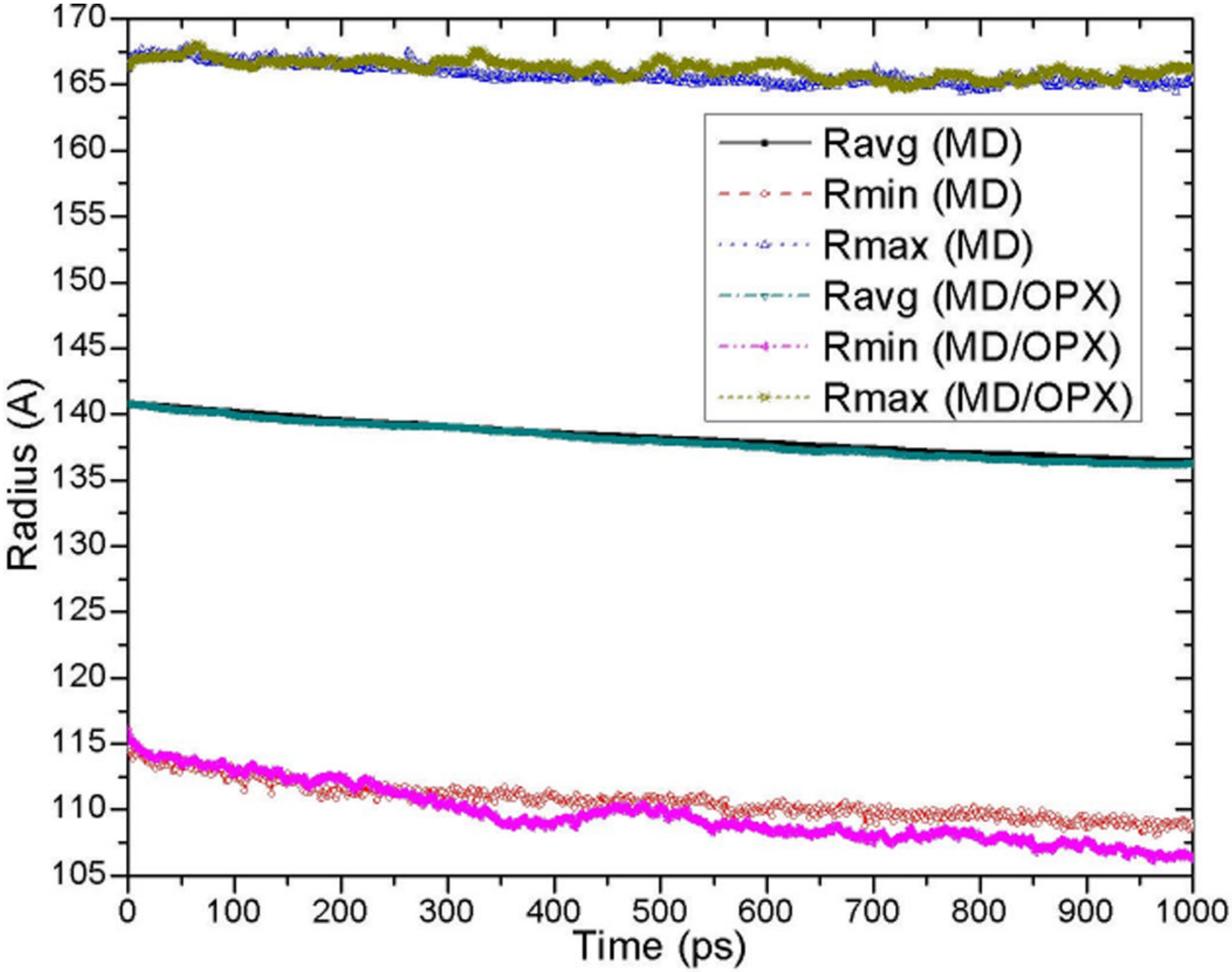

Radius (A)
Time (ps)
Ravg (MD)
Rmin (MD)
Rmax (MD)
Ravg (MD/OPX)
Rmin (MD/OPX)
Rmax (MD/OPX)
170
165
160
155
150
145
140
135
130
125
120
115
110
105
0
100
200
300
400
500
600
700
800
900
1000

Fig. 6.

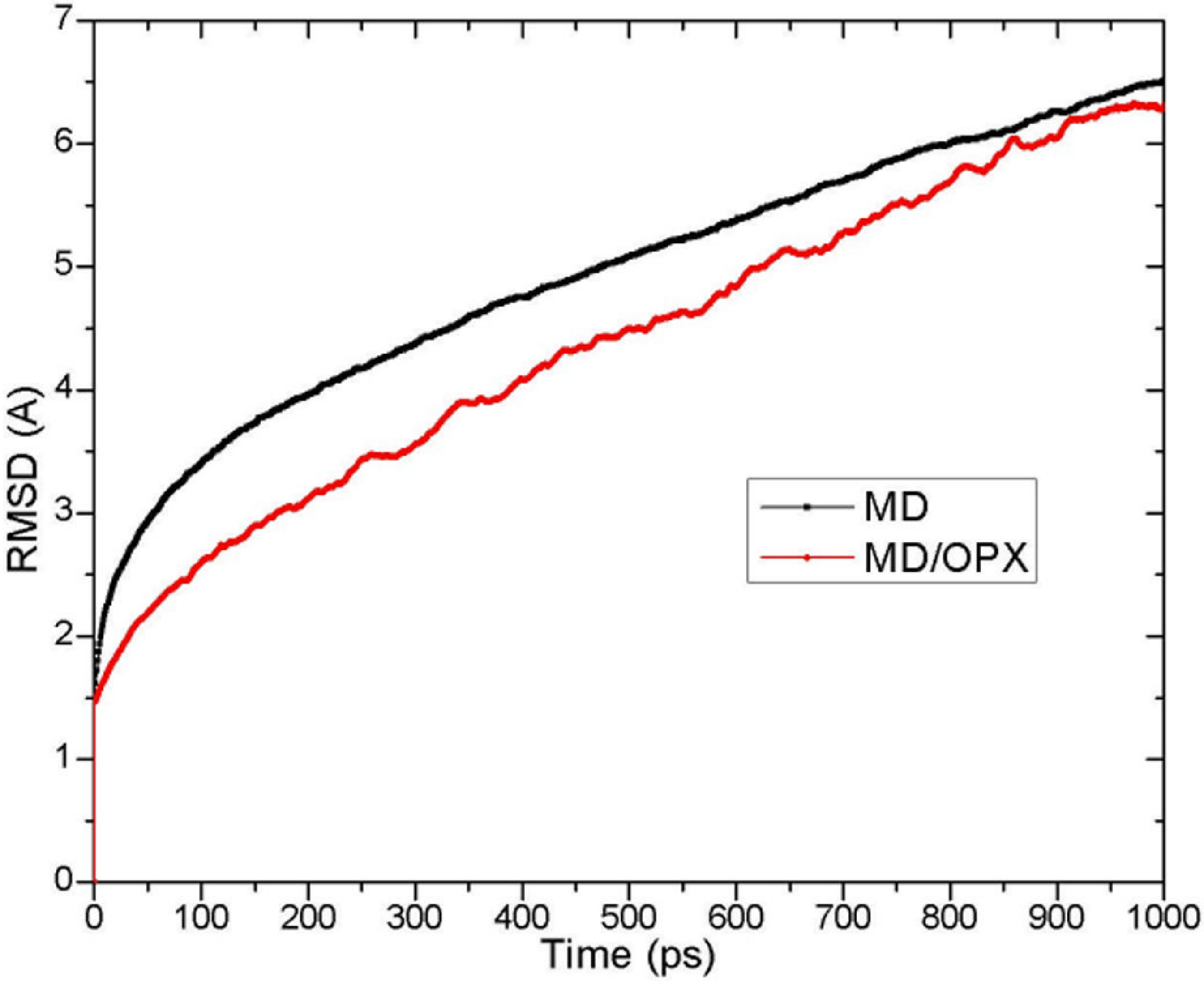
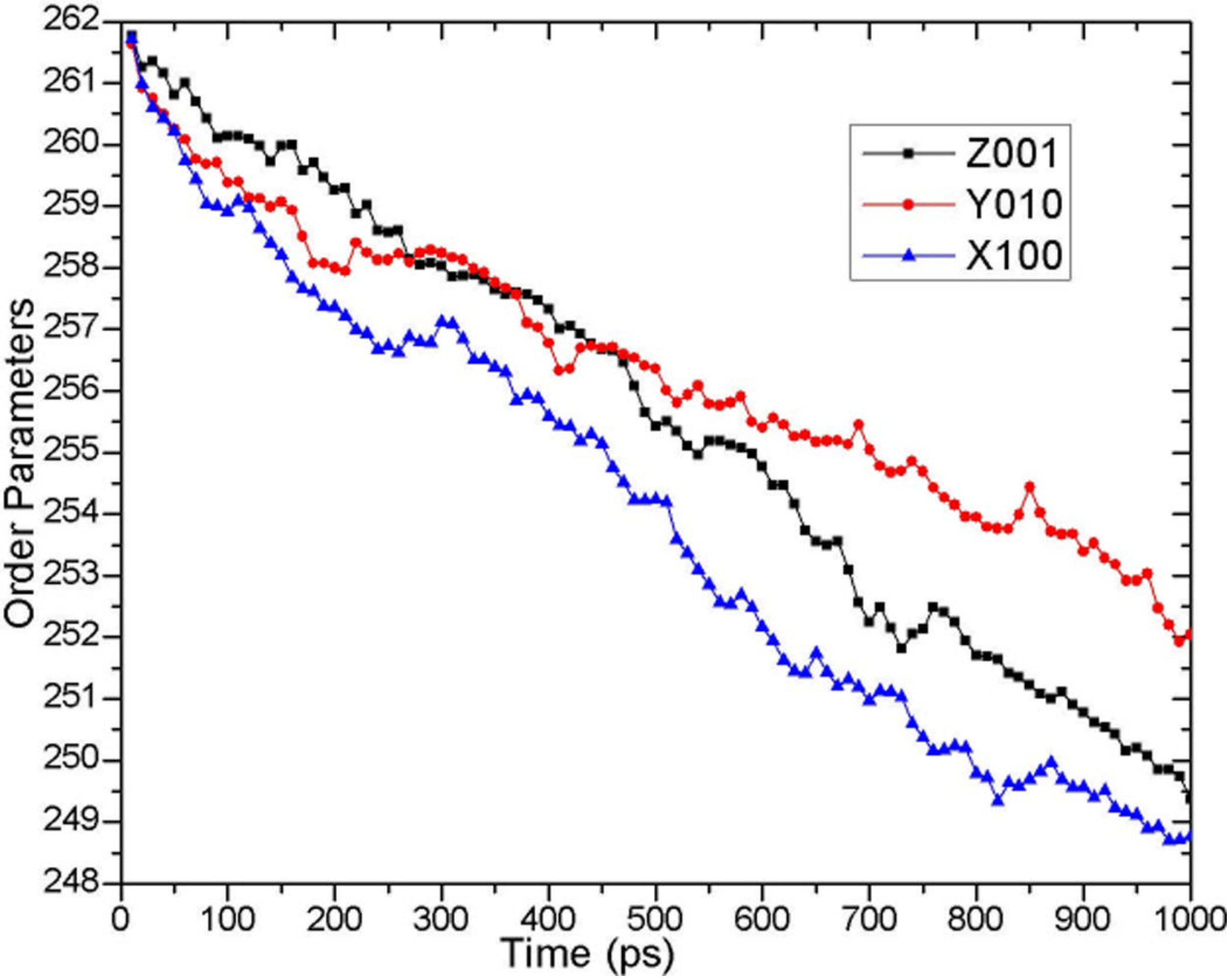
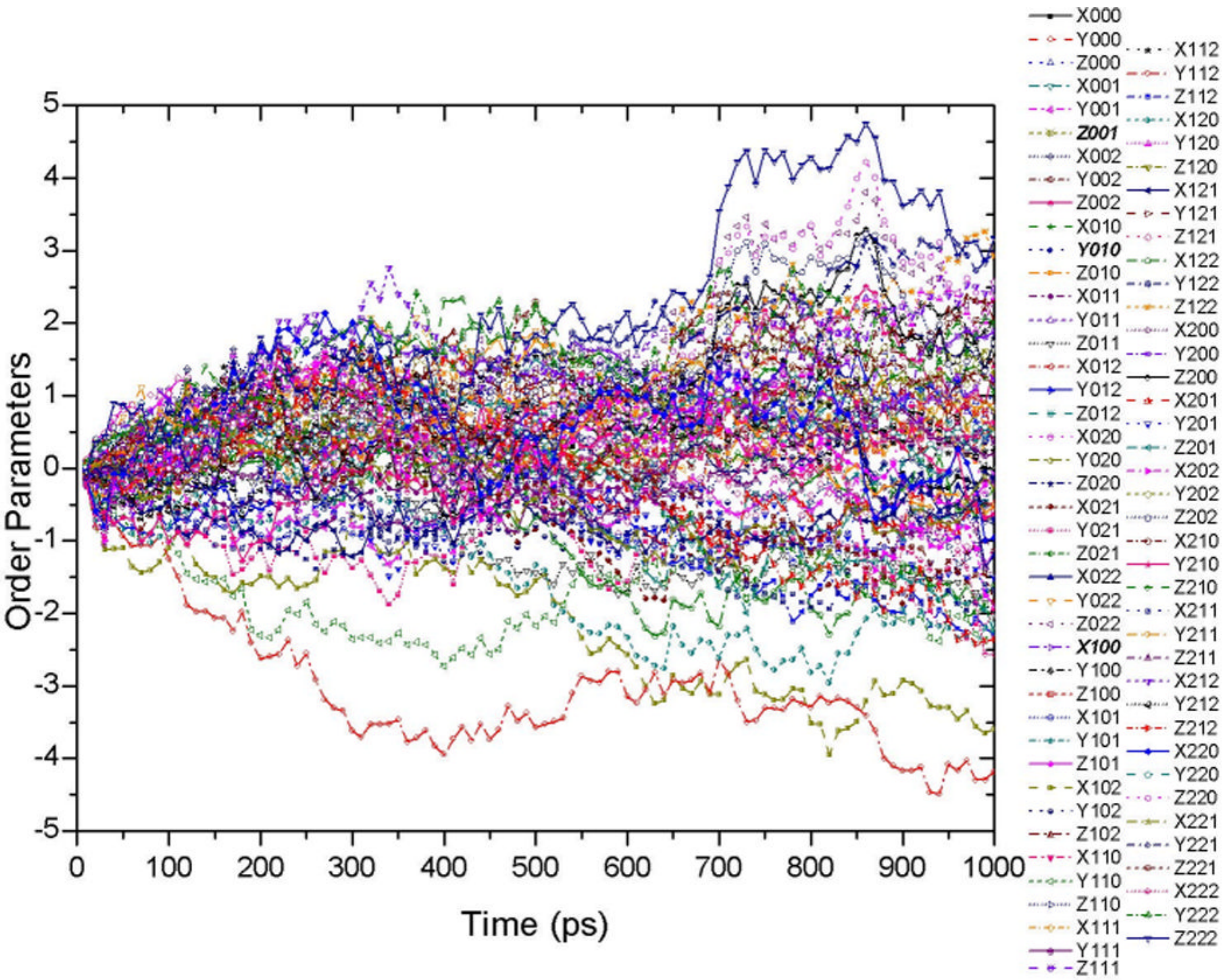

RMSD (A)
0
1
2
3
4
5
6
7
Time (ps)
0
100
200
300
400
500
600
700
800
900
1000
MD
MD/OPX

Fig. 7.


Order Parameters
262
261
260
259
258
257
256
255
254
253
252
251
250
249
248
0
100
200
300
400
500
600
700
800
900
1000
Time (ps)
Z001
Y010
X100


Order Parameters
5
4
3
2
1
0.0
-1
-2
-3
-4
-5
Time (ps)
0
100
200
300
400
500
600
700
800
900
1000
X000
Y000
Z000
X001
Y001
Z001
X002
Y002
Z002
X010
Y010
Z010
X011
Y011
Z011
X012
Y012
Z012
X020
Y020
Z020
X021
Y021
Z021
X022
Y022
Z022
X100
Y100
Z100
X101
Y101
Z101
X102
Y102
Z102
X110
Y110
Z110
X111
Y111
Z111
X112
Y112
Z112
X120
Y120
Z120
X121
Y121
Z121
X122
Y122
Z122
X200
Y200
Z200
X201
Y201
Z201
X202
Y202
Z202
X210
Y210
Z210
X211
Y211
Z211
X212
Y212
Z212
X220
Y220
Z220
X221
Y221
Z221
X222
Y222
Z222

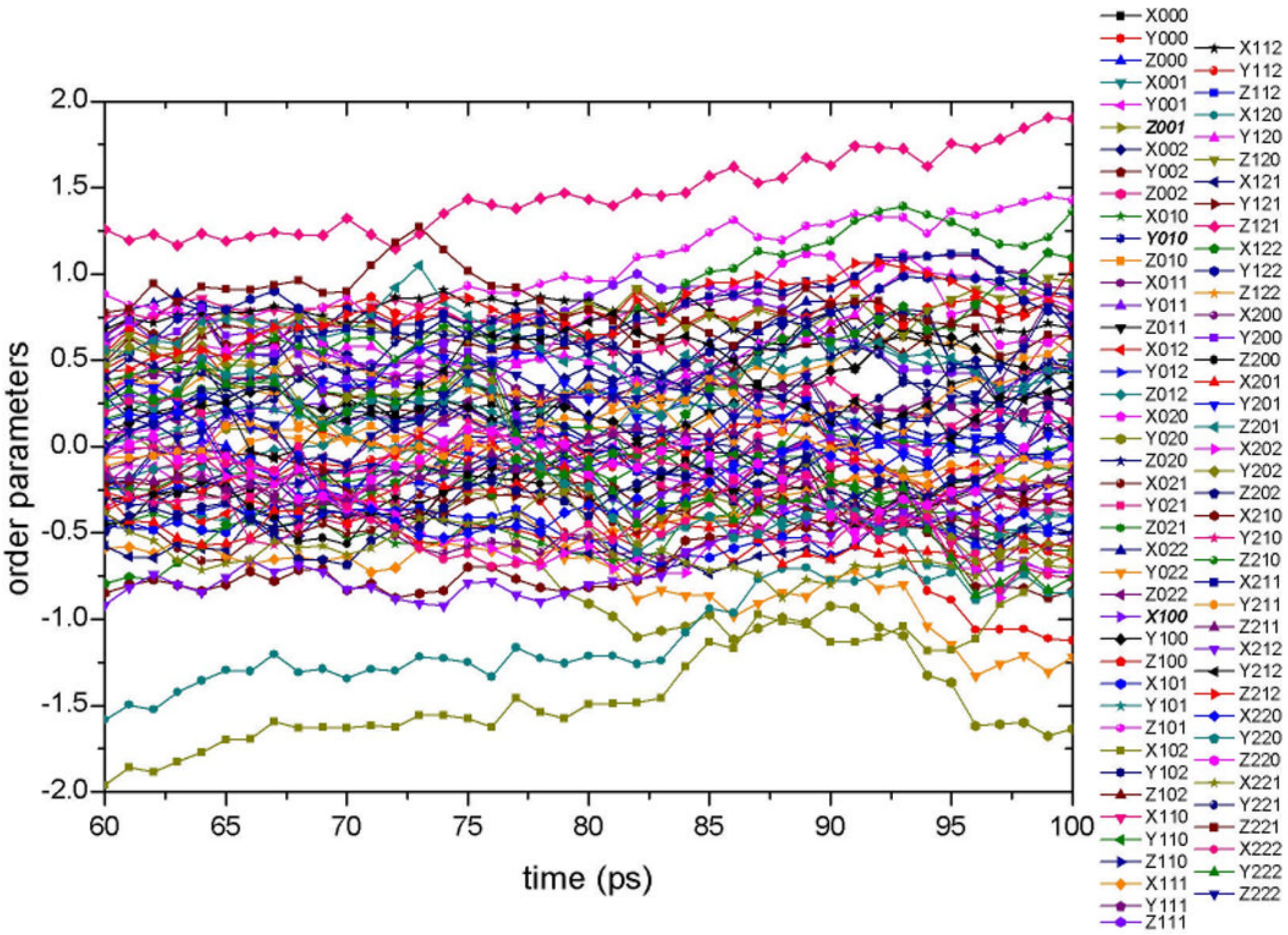

2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
-2.0
order parameters
60
65
70
75
80
85
90
95
100
time (ps)
X000
Y000
Z000
X001
Y001
Z001
X002
Y002
Z002
X010
Y010
Z010
X011
Y011
Z011
X012
Y012
Z012
X020
Y020
Z020
X021
Y021
Z021
X022
Y022
Z022
X100
Y100
Z100
X101
Y101
Z101
X102
Y102
Z102
X110
Y110
Z110
X111
Y111
Z111
X112
Y112
Z112
X120
Y120
Z120
X121
Y121
Z121
X122
Y122
Z122
X200
Y200
Z200
X201
Y201
Z201
X202
Y202
Z202
X210
Y210
Z210
X211
Y211
Z211
X212
Y212
Z212
X220
Y220
Z220
X221
Y221
Z221
X222
Y222
Z222

Fig. 8.

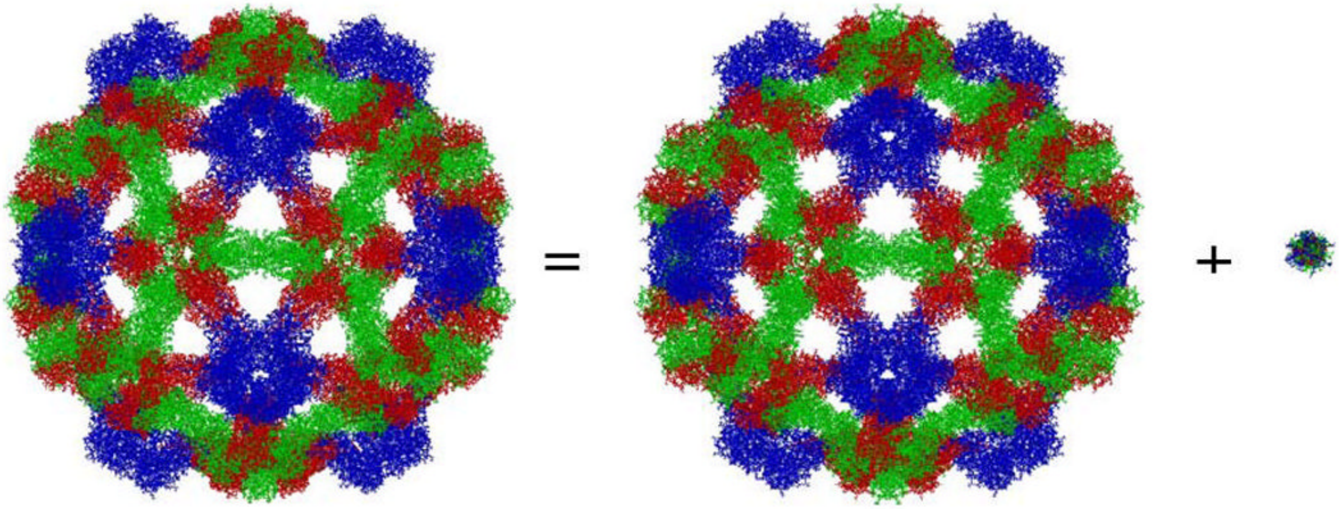


Fig. 9.

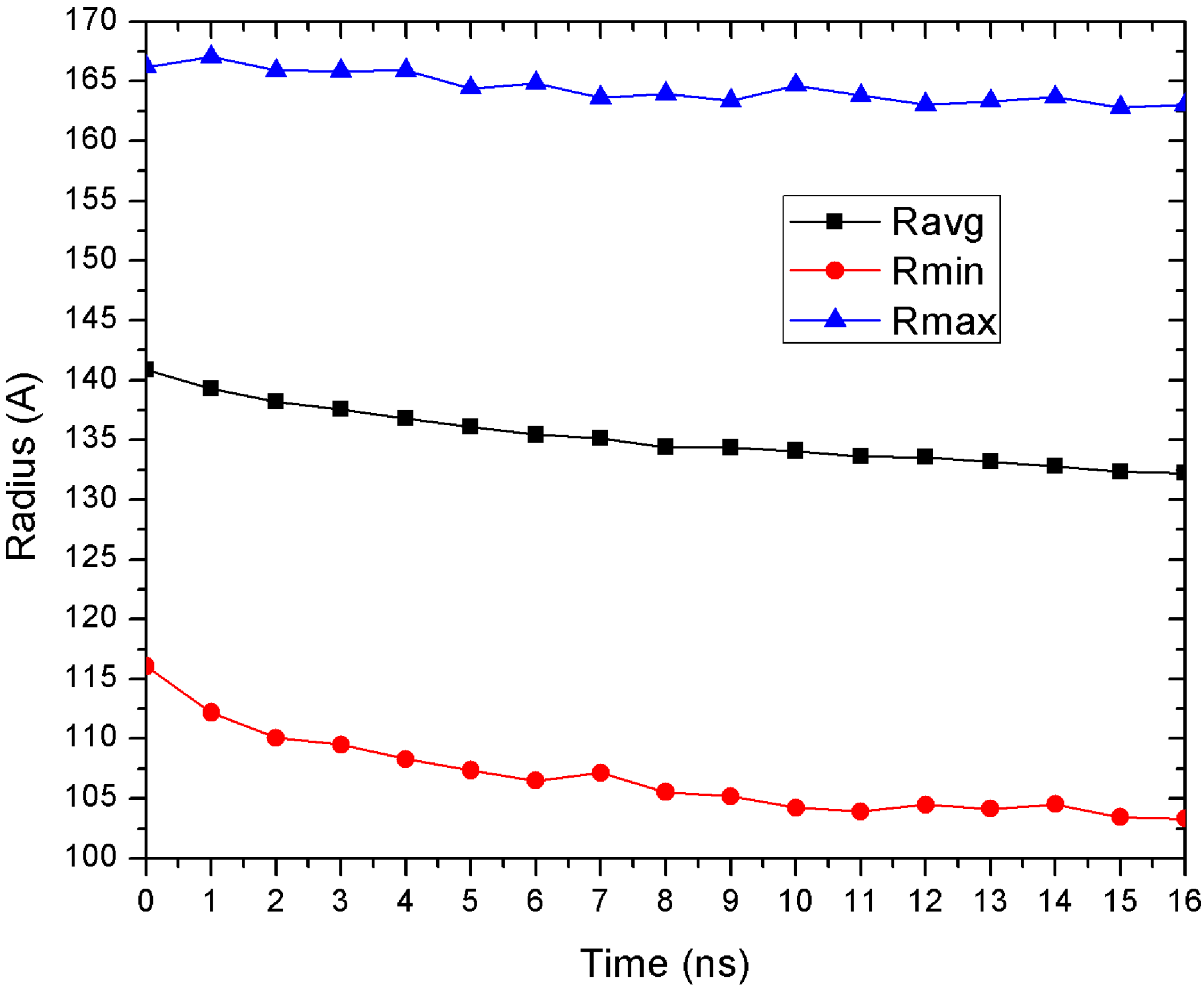

Radius (A)
Time (ns)
Ravg
Rmin
Rmax
100
105
110
115
120
125
130
135
140
145
150
155
160
165
170
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16

Fig. 10.