

# ALL-ATOM MULTISCALE COMPUTATIONAL MODELING OF VIRAL DYNAMICS

Yinglong Miao

Submitted to the faculty of University Graduate School

In partial fulfillment of the requirements

For the degree

Doctor of Philosophy

In the Department of Chemistry,

Indiana University

August 2009

Accepted by the Graduate Faculty, Indiana University, in partial fulfillment of the requirements  
for the degree of Doctor of Philosophy.

Doctoral Committee

---

Peter J. Ortoleva, Ph.D.

---

Krishnan Raghavachari, Ph.D.

---

Srinivasan S. Iyengar, Ph.D.

Date of Dissertation Defense  
August 17, 2009

---

David P. Giedroc, Ph.D.

© 2009

Yinglong Miao

ALL RIGHTS RESERVED

To my beloved parents, my two dearest brothers and their families.

## Acknowledgements

I would like to express my sincerest gratitude and thanks to my advisor Professor Peter J. Ortoleva for guiding me through my graduate study and research at the Center for Cell and Virus Theory in the chemistry department of Indiana University. His patience and strong support greatly helped me adapting myself to graduate course studies and research in theoretical chemistry. His guidance enabled me to develop new theoretical methods to address complex problems in theoretical chemistry and construct computational models for challenging simulation of large biophysical nanosystems. His thoughtful suggestions and comments shaped me to be a more professional scientist for pursuing exciting research projects with excellence and accuracy. I would also like to thank members of my graduate committee, Professor Krishnan Raghavachari, Professor Srinivasan S. Iyengar, Professor David P. Giedroc and Professor Martin Stone, for their valuable discussions, suggestions and comments.

I would pass my heartfelt thanks to my parents and my two brothers' families. Without their constant support and help all through the years, I would not be able to focus on my graduate study and finish my Ph.D. program successfully. I would also pass my deep appreciation to my friends, especially Bo Liang, Haitao Wang and Yi Zeng, for their help and valuable conversations clarifying my confusions in life and research. My final thanks go to all group members at the Center for Cell and Virus Theory, active hiking participants of the Bloomington Hiking Group, and my fellow students in Indiana University for their kind help and support.

# Abstract

Yinglong Miao

## All-atom Multiscale Computational Modeling of Viral Dynamics

Viruses are composed of millions of atoms functioning on supra-nanometer length scales over timescales of milliseconds or greater. In contrast, individual atoms interact on scales of angstroms and femtoseconds. Thus they display dual microscopic/macrosopic characteristics involving processes that span across widely-separated time and length scales. To address this challenge, we introduced automatically generated collective modes and order parameters to capture viral large-scale low-frequency coherent motions. With an all-atom multiscale analysis (AMA) of the Liouville equation, a stochastic (Fokker-Planck or Smoluchowski) equation and equivalent Langevin equations are derived for the order parameters. They are shown to evolve on timescales much larger than the  $10^{-14}$ -second timescale of fast atomistic vibrations and collisions. This justifies a novel multiscale Molecular Dynamics/Order Parameter eXtrapolation (MD/OPX) approach, which propagates viral atomistic and nanoscale dynamics simultaneously by solving the Langevin equations of order parameters implicitly without the need to construct thermal-average forces and friction/diffusion coefficients. In MD/OPX, a set of short replica MD runs with random atomic velocity initializations estimate the ensemble average rate of change in order parameters, extrapolation of which is then used to project the system over long time. The approach was implemented by using NAMD as the MD platform. Application of

MD/OPX to cowpea chlorotic mottle virus (CCMV) capsid revealed that its swollen state undergoes significant energy-driven shrinkage in vacuum during 200ns simulation, while for the native state as solvated in a host medium at pH 7.0 and ionic strength  $I=0.2M$ , the N-terminal arms of capsid proteins are shown to be highly dynamic and their fast fluctuations trigger global expansion of the capsid. Viral structural transitions associated with both processes are symmetry-breaking involving local initiation and front propagation. MD/OPX accelerates MD for long-time simulation of viruses, as well as other large bionanosystems. By using universal inter-atomic force fields, it is generally applicable to all dynamical nanostructures and avoids the need of parameter recalibration with each new application. With our AMA method and MD/OPX, viral dynamics are predicted from laws of molecular physics via rigorous statistical mechanics.

---

---

---

---

# Table of Contents

Acknowledgements .....	v
Abstract .....	vi
Table of Contents .....	viii
List of Figures .....	xii
List of Tables.....	xix
 Part I. Background .....	 1
Overview.....	1
Chapter 1 Structures and Dynamics of Viruses .....	3
1.1 Virus Structures .....	3
1.2 Viral Dynamics .....	8
Chapter 2 Viral Computational Modeling.....	14
2.1 Review of Viral Computational Models.....	14
2.2 Multiscale Analysis.....	18
2.3 All-atom Multiscale Computational Modeling.....	22
 Part II. Methods.....	 25
Overview.....	25
Chapter 3 All-atom Multiscale Analysis (AMA).....	26



3.1	AMA of the $N$ -atom Liouville Equation.....	26
3.2	Nanocanonical Ensemble Approach.....	31
3.3	Derivation of a Fokker-Planck equation for Stochastic Dynamics of Order Parameters.....	35
3.4	Collective Modes and Order Parameters for $N$ -atom Nanosystems .....	40
3.5	Free energy and friction coefficients .....	46
3.6	Generalizations .....	48
Chapter 4	Transition State Theory for Viral Dilatational Structural Transitions .....	50
4.1	Multiscale Viral Dilational Structural Transition Model .....	50
4.2	Order Parameter Dependences of System Energy and Lowest Order Probability Density .....	55
4.3	Derivation of the Fokker-Planck equation of Order Parameters .....	57
4.4	Transition State Theory .....	63
Chapter 5	Molecular Dynamics/Order Parameter eXtrapolation (MD/OPX) .....	67
5.1	Characteristics of Order Parameters .....	67
5.2	MD/OPX Approach .....	71
5.3	MD/OPX Implementation and Optimization .....	74
Part III.	Results and Discussion.....	79
	Overview.....	79
Chapter 6	Application of Transition State Theory to NoV Capsid .....	80
6.1	Hysteresis, Irreversibility and the Free Energy Profile in Viral Structural Transitions.....	80

6.2	Free Energy Barrier of NøV capsid Structural Transition .....	82
Chapter 7 MD/OPX Validation with CCMV Capsid Structural Transitions .....		86
7.1	Background of CCMV and its Capsid Structural Transitions .....	86
7.2	Order Parameters Capturing Capsomer Dynamics during CCMV Capsid Expansion	88
7.3	Comparison of MD and MD/OPX Simulations on CCMV Capsid.....	93
7.4	MD/OPX Performance .....	96
Chapter 8 Viral Capsid Structural Transition Mechanisms Revealed via Long-time MD/OPX Simulation .....		98
8.1	Insights into CCMV Capsid Stability through Short-Time MD Simulations.....	98
8.2	Shrinkage of Swollen CCMV Capsid in vacuum Captured by Long-time MD/OPX Simulation.....	100
8.3	Mechanisms of Viral Capsid Structural Transition.....	107
8.4	Shrinkage of Swollen CCMV Capsid is Energy-driven .....	113
Chapter 9 All-atom Multiscale MD/OPX Simulation of CCMV Capsid Swelling .....		117
9.1	Insights into CCMV Capsid Swelling through Short-Time MD Simulations .....	117
9.2	MD/OPX Simulation of CCMV Capsid Swelling.....	121
9.3	CCMV Capsid Swelling Mechanisms .....	126
9.4	Dependence of CCMV Capsid Swelling on System Conditions.....	130
9.5	Prospects of multiscale MD/OPX simulation.....	136
Part IV. Summary and Prospects .....		139

Appendix A: Decomposition of the Liouville equation .....	146
Appendix B: Fourier/bilateral Laplace transformation .....	149
Bibliography .....	150

## List of Figures

**Figure I.1** Icosahedral “quasi-equivalence” geometric model: (a) an icosahedral viral capsid can be constructed by replacing 12 hexamers with pentamers at appropriate positions in a net of hexamers, in which the origin and the  $(h, k)$  hexamers are replaced by pentamers with  $h$  and  $k$  as zero or positive integers. A triangulation number  $T$  is defined as  $T=h^2+hk+k^2$  to classify the icosahedral quasi-equivalent lattices with  $T$  equal to 1, 3, 4, 7, etc. (b) the construction of a  $T=3$  ( $h=1, k=1$ ) icosahedron: a  $T=3$  icosahedral face is shaded in a net of seven hexamer units and twenty such identical faces can be arranged and folded into a three-dimensional quasi-equivalent icosahedral lattice.

**Figure I.2** Four quasi-equivalent icosahedral viral capsids: (a)  $T=3$  CCMV ( $h=1, k=1$ ; PDB ID: 1CWP), (b) pseudo  $T=3$  poliovirus ( $h=1, k=1$ ; PDB ID: 2PLV), (c)  $T=4$  N $\omega$ V ( $h=2, k=0$  or  $h=0, k=2$ ; PDB ID: 1OHF), and (d)  $T=7$  HK97 bacteriophage ( $h=2, k=1$  or  $h=1, k=2$ ; PDB ID: 1OHG) (images downloaded from VIPERdb website: <http://viperdbscripps.edu/>).

**Figure I.3** Native CCMV swells about 10% as pH is increased from 5.0 to 7.5 in the absence of divalent cations.

**Figure I.4** Order parameters characterizing nanoscale features affect the relative probability of atomistic configurations which, in turn, mediates the forces driving order parameter dynamics.

**Figure I.5** The essence physical picture of a virus: a massive and geometrically large aggregate of atoms immersed in a host medium that subjects it to frequent collisions.

**Figure II.1** (a) a double well free energy profile  $F$  vs.  $\Phi$  implied by the hysteresis in N $\omega$ V capsid structural transition and (b) the associated probability distribution  $W$  vs.  $\Phi$ .

**Figure II.2** Time courses of CCMV capsid order parameters corresponding to Legendre polynomials in the X, Y, and Z directions of the MD simulation: (a) three order parameters 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 can be readily extrapolated to long time on a timescale of ns, (b) other order parameters fluctuating rapidly around zero over 1ns and (c) a closer look at them from 60ps to 100ps showing that their characteristic time is roughly 10ps.

**Figure II.3** A snapshot of the CCMV capsid after 1ps MD simulation showing that the atomic coordinates are equal to their coherent contribution plus the residuals.

**Figure II.4** Schematic flowchart of MD/OPX (a) and its implementation with dynamic adaptive run parameter adjustment (b): energy minimization and a short  $\delta t_0$  MD run are applied to anneal the simulated all-atom structure after OPX and run parameters (notably  $\delta t$  and  $\Delta t$ ) are adjusted dynamically according to structural and dynamic indicators to optimize the balance between program accuracy and CPU speed.

**Figure II.5** Schematic flowchart of MD/OPX for simulation of nanostructures solvated in host media with modules for “solvation”, “extraction of nanostructure snapshots” and “resolution” integrated.

**Figure III.1** The cleavage-defective mutant of N $\omega$ V undergoes hysteretic structural transition when pH in the host medium changes. The curve indicates values of  $\Phi$  for which the virus/host system’s free energy is a minimum.

**Figure III.2** (a) Crystal structure of a CCMV protomer with its three quasi-threefold related subunits colored in blue (A), red (B) and green (C). (b) Native CCMV capsid organized in 12 pentameric and 20 hexameric capsomers with 5 blue A subunits in each pentamer and 3 red B and 3 green C subunits in each hexamer. (c) The swollen CCMV capsid generated computationally by rigid-body translations and rotations of the pentamers and hexamers according to the expansion scheme proposed by Liu et al.[1].

**Figure III.3** (a1) to (a5) in the left column show the inner shell view of the 4 intermediate CCMV capsid structures and the final swollen one generated from the native capsid and (b1) to (b5) in the right column show their corresponding coherent structures calculated from  $3^3$  order parameters.

**Figure III.4** Demonstration of using order parameters to capture the translation and rotation of a pentamer (a) and a hexamer (b) during the native CCMV capsid expansion: the black bead represents the center-of-mass (COM) position of the capsid, which stays at the origin for all structures; blue beads represent COM positions of the protein subunits in the chosen pentamer and hexamer from the 5 original capsid structures; while the red, green, cyan and purple beads correspond to the protein subunits from the 5 coherent capsid structures calculated from  $3^3$ ,  $4^3$ ,  $5^3$  and  $6^3$  order parameters.

**Figure III.5** Ribbon representations of the output structures of the swollen CCMV capsid after 1ns simulation using MD (a) and MD/OPX (b).

**Figure III.6** Comparison of 1ns MD and MD/OPX simulations: (a) Time courses of the average, minimum and maximum radii of the CCMV capsid backbone (b) RMSD of atomic

positions between the simulated CCMV capsid structure along the trajectory and its starting structure.

**Figure III.7** 10ns MD simulation of the native and swollen states of CCMV capsid: (a) time courses of the decrement in capsid average radius from their initial structures and (b) RMSD of atomic positions between their trajectory snapshots and the initial structures.

**Figure III.8** 200ns MD/OPX simulation of the swollen CCMV capsid: (a) time courses of the decrement in the average, minimum and maximum radii of the capsid backbone from the simulation starting structure, and (b) RMSD of atomic positions between capsid snapshots along the trajectory and the initial structure.

**Figure III.9** Interior views of (a) swollen CCMV capsid, (b) the final structure of 200ns MD/OPX simulation, and (c) the COM displacements of protein subunits from their initial configurations to the final that depict shrinkage of the capsid.

**Figure III.10** (a) RMSD of atomic positions for a selected pentamer and a hexamer (see Figure 7a for P1 and H1) between trajectory snapshots and their initial configurations in comparison to that for the entire capsid, (b) the average COM translation distance as a function of time for pentamers and hexamers, and (c) time courses of the average rotation angle for pentamers and hexamers calculated through fitting their structures to the initial configurations.

**Figure III.11** (a) Schematic representation of icosahedral CCMV capsid with 12 pentamers (blue pentagons) labeled from P1 to P12 and 20 hexamers (red hexagons) from H1 to H20, and time courses of the COM translation distance of (b) pentamers and (c) hexamers from their original positions.

**Figure III.12** Interior view of the capsid shrinkage trajectory snapshots with atoms colored by their displacements from the original positions in a BWR color scale (0Å for blue, 24Å for red, and white as the midpoint): (a) initial state with all atoms in blue, (b) 5ns, (c) 20ns, (d) 50ns, (e) 100ns, and (f) 200ns of simulation. It can be seen that atoms in the upper pentamer of shown trajectory snapshots (i.e., P11 as labeled in Figure III.11a) start to change their color from blue to white and to red first, and then this color change propagates across the capsid. Thus capsid shrinkage is a symmetry-breaking/front propagation process.

**Figure III.13** Time courses of RMSD of atomic positions between pentamer P1 as labeled in Figure III.11a and the 11 others after rigid-body transformation of P1 to the positions of other pentamers according to icosahedral symmetry.

**Figure III.14** Variations of the system temperature (a) and energies (b) in one typical MD/OPX cycle capturing 3,010,100-3,075,100fs of the simulation, and (c) the system kinetic and total energies plotted versus time during 200ns capsid shrinkage.

**Figure III.15** X-ray crystal structure of CCMV protomer with the missing N-terminal residues added and optimized. Three quasi-threefold related protein subunits are colored in blue (A), red (B) and green (C) respectively, and the N-terminal arms in yellow.

**Figure III.16** Schematic representations of CCMV capsid solvated in (a) a water cube and (b) a water truncated octahedron at pH 7.0 and ionic strength I=0.2M used for our simulations with the water box created using VMD “volmap” plugin and ions represented in CPK.

**Figure III.17** 3ns MD simulation of native CCMV capsid solvated in a water cube with an initial 7Å water boundary: (a) changes in the average, minimum and maximum radii of the



capsid backbone from the starting configuration, (b) averaged thickness of water layer surrounding the capsid backbone in X, Y and Z directions, and (c) RMSD of atomic positions between snapshots of capsid backbone along the trajectory and the starting configuration.

**Figure III.18** MD/OPX trajectory snapshots showing the expansion of native CCMV capsid (Ncap) in water cube with an initial 15Å water boundary: (a) the starting configuration, (b) 0.5ns, (c) 1.5ns and (d) 3ns.

**Figure III.19** Analysis of 3ns MD/OPX trajectory of native CCMV capsid (Ncap): (a) time course of the change in capsid average radius from the starting configuration, (b) RMSD of atomic positions between snapshots of capsid backbone along the trajectory and the starting configuration, and (c) averaged thickness of water boundary surrounding the capsid backbone in X, Y and Z directions.

**Figure III.20** 17.04ns MD/OPX simulation of the swelling of capsid intermediate (Icap): (a) time course of the change in the capsid average radius and (b) RMSD of atomic positions between snapshots of capsid backbone along the trajectory and the initial structure.

**Figure III.21** Analysis of CCMV capsid along the entire 18.54ns MD/OPX trajectory: (a) the volume of the capsid (i.e., protein shell) and its cavity, and (b) the capsid SASA.

**Figure III.22** Collective motions of pentamers and hexamers during 18.54ns CCMV capsid swelling: time courses of (a) their average COM translation distances and (b) their average rotation angles calculated through structure fitting to the initial configurations in native CCMV capsid.

**Figure III.23** Interior view of the capsid swelling trajectory snapshots with atoms colored by their displacements from the original positions in a BWR color scale (0Å for blue, 24Å for red, and white as the midpoint): (a) initial native state with all atoms in blue, (b) 1.8ns, (c) 3ns, (d) 6ns, (e) 12ns, and (f) 18ns. N-terminal arms of the capsid proteins are highlighted in ribbons and they are colored according to the displacements of atoms as well. It is seen that the N-terminal arms change their color from blue to red during early stage of the simulation and their structural changes trigger global expansion of the capsid. Swelling of CCMV capsid is a symmetry-breaking process involving local initiation and front propagation.

## List of Tables

**Table III.1** Simulations showing CCMV capsid swelling is strongly dependent on system conditions: <sup>a</sup> Capsid interface is the minimum distance between water molecules and atoms on the capsid surface used for solvation, <sup>b</sup> water boundary is the minimum distance between atoms on the capsid surface and faces of the water box (square face for cube and hexagonal face for truncated octahedron); <sup>c</sup> there are 522,000 atoms in CCMV capsid and the total number of atoms in the systems ranges from ~2.41 to ~2.86 million depending on the shape of water box used for solvation and the thickness of water boundary; <sup>d</sup> NPT is constant pressure (1atm) and temperature (298.15K) ensemble and NVT is constant volume (unchanged from the initial dimensions of systems prepared with VMD) and temperature (298.15K) ensemble; <sup>e</sup> performance here is for running simulations with 256 processors on Indiana University Big Red cluster; and <sup>f</sup> the capsid radius calculated is the average distance of non-hydrogen atoms in the capsid backbone to their center of mass.

## **Part I. Background**

### **Overview**

Viruses are small infectious agents that can pose global threats to human health [2]. Such human pathogens include H1N1 influenza virus, SARS coronavirus, poliovirus, dengue virus, Human Papilloma virus and Human Rhinovirus. They cause a wide variety of diseases including influenza flu, poliomyelitis, AIDS, etc. To address viral threats necessitates the understanding of their structures and infection processes, e.g., assembly, structural transitions and disassembly. For rapid responses and pre-emptive strategies, forward-looking approaches are needed to predict properties of emerging viral strains before they become a pandemic. While mutants of viruses could be generated and analyzed experimentally as part of a forecasting strategy by using X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, cryo-electron microscopy (cryoEM), solution X-ray scattering and biochemical techniques [3, 4], the potential risks and time bottlenecks are unacceptable.

With rapid development of computational power and advances in molecular modeling[5], viruses can be easily mutated and computational modeling of virus or virus-like particles could provide a cost-effective approach for understanding their properties and thus facilitate the design of anti-viral drugs and vaccines[6]. In order to evaluate the interactions of a virus with therapeutical small molecules or a cell surface receptor[7, 8], a computational model that accounts for viral atomistic details is needed. Also, the prediction of whole virus behavior is

required, e.g., the evaluation of vaccine stability and the responses of viruses to anti-viral drugs. However, viruses are composed of millions of atoms functioning on supra-nanometer length scales over timescales of milliseconds or greater, while individual atoms interact on scales of angstroms and femtoseconds [9, 10]. As a result, the understanding of viral dynamics spanning widely separated time and length scales presents challenge in biophysical chemistry, structural virology and nano-medicine.

Here an all-atom multiscale approach that addresses atomistic fluctuations and nanoscale processes simultaneously is presented for studying viral dynamics. It includes a theoretical all-atom multiscale formulation for viruses (i.e., all-atom multiscale analysis of the  $N$ -atom Liouville equation) [9, 11] and the implementation of a multiscale computational model for simulating viral dynamics, i.e., Molecular Dynamics/Order Parameter eXtrapolation (MD/OPX) [12-14]. With our unique AMA and MD/OPX approach, principles of viral dynamics are derived from laws of molecular physics via rigorous statistical mechanics and mathematical derivations. Both the mathematics and physics show that viruses undergo nanoscale coherent structural changes that involve collective motions of atoms. By capturing such phenomena with our automatically constructed collective modes and order parameters, viral dynamics can be effectively described and their modeling can be greatly facilitated. The underlying all-atom description used in our multiscale approach enables us to study a wide range of other bionanosystems as well, e.g., proteins, ribosomes and liposomes.

## **Chapter 1 Structures and Dynamics of Viruses**

### **1.1 Virus Structures**

Viruses are amongst the simplest life forms. Their structural descriptions are important to pure and applied studies of viral properties and infection mechanisms. They serve as an ideal system for validating our computational modeling approach. Viral behaviors revealed through our simulations could facilitate the design of anti-viral drugs and vaccines, functionalized nanoparticles for medical imaging and thermal cancer treatments, and nanocapsules for delivery of therapeutic agents in health sciences and biotechnology.

Viruses exist in a wide range of sizes, shapes and internal structures from filamentous, rod-shaped to spherical, enveloped or non-enveloped and contain double-stranded DNA, single-stranded DNA, or single-stranded RNA as their genetic material [2]. Here we restrict the discussion to non-enveloped spherical RNA viruses exhibiting icosahedral symmetry, which constitutes the most studied group of viruses. They are basically composed of an outer protein shell, referred to as capsid, and RNA nucleic acid as genetic material. The primary role of the viral capsid is to protect the enclosed genetic material against adverse conditions and processes existing inside a cell or extracellular medium (e.g., enzymatic digestion).

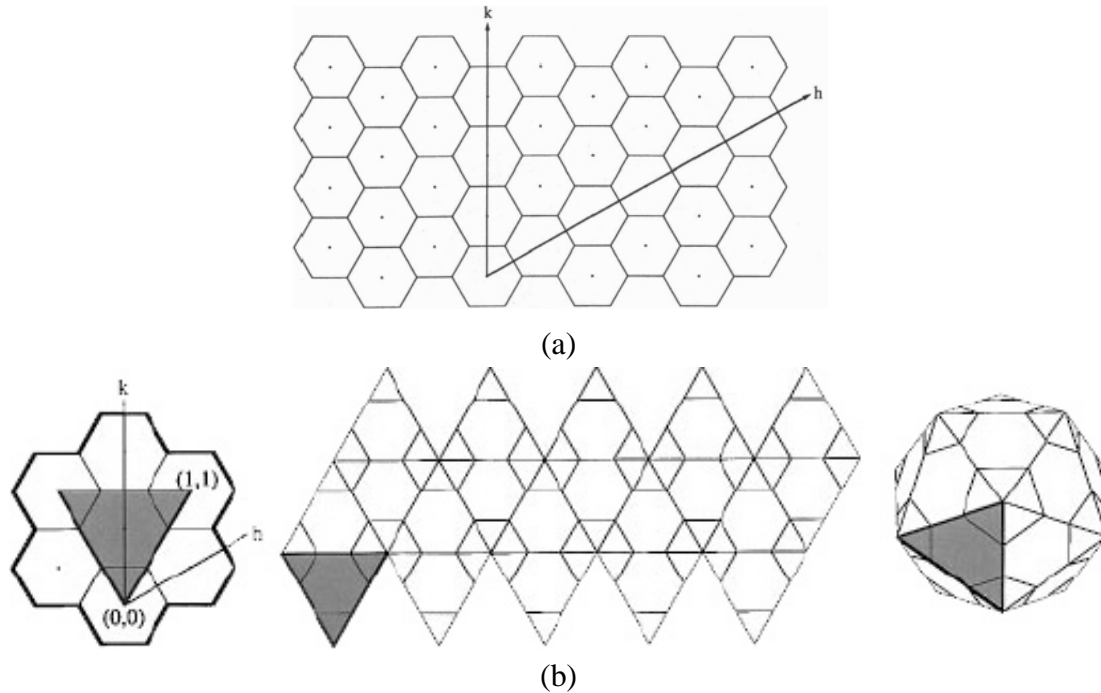
The genetic efficiency of a virus requires that its capsid assembles from multiple copies of one or few types of proteins instead of from a large number of different proteins. The simplest icosahedral virus has 60 copies of a protein subunit in its capsid with each located in structural identical environments on the capsid surface. As viral capsid increases its size for larger volume

to package more genetic material, more than one protein can be arranged in its asymmetric unit (referred to as capsid protomer), and 60 copies of the asymmetric unit will assemble to form the icosahedral capsid.

In the “quasi-equivalence” geometric model proposed by Caspar and Klug [15, 16], an icosahedral viral capsid, composed of pentamers and hexamers of protein subunits, is generated by replacing 12 hexamers with pentamers at appropriate positions in a net of hexamers (Figure I.1a). Hexamers are initially considered planar and pentamers are considered convex, introducing curvature in the sheet of hexamers when they are inserted. To construct a model for an icosahedral quasi-equivalent lattice, one face of the icosahedron (i.e., an equilateral triangle connecting three pentamers) is first generated in a net of hexamers by replacing three of them with pentamers at the following positions: the origin, the  $(h, k)$  hexamer with  $h$  and  $k$  as zero or positive integers, and the third identified to complete the equilateral triangle. Twenty identical copies of the icosahedral face can then be arranged properly and folded into a three-dimensional quasi-equivalent lattice. A triangulation number  $T$  is defined as  $T=h^2+hk+k^2$  to classify icosahedral quasi-equivalent lattices. Typically,  $T$  is equal to 1, 3, 4, 7, etc. and could be as large as 25.

An example showing the construction of a  $T=3$  ( $h=1, k=1$ ) icosahedral shell is shown in Figure I.1b. The icosahedral face is defined by a triangle (see the shaded area) in a net of seven hexamer units with hexamers at the origin  $(0, 0)$ ,  $(1, 1)$  and  $(-1, 1)$  replaced by pentamers. The capsid asymmetric unit is one-third of the icosahedral face. It contains three quasi-equivalent subunits (two from the hexamer in the middle of the face and one from a pentamer). Twenty

identical copies of the icosahedral face are arranged as in the middle chart of Figure I.1b, and can be folded into the three-dimensional icosahedral shell as shown. Quasi-equivalent icosahedral capsids normally contain  $60T$  protein subunits, which are arranged into 12 pentamers and  $10*(T-1)$  hexamers.



**Figure I.1** Icosahedral “quasi-equivalence” geometric model: (a) an icosahedral shell can be constructed by replacing 12 hexamers with pentamers at appropriate positions in a net of hexamers. One face of the icosahedral shell can be generated by replacing hexamers at the origin and  $(h, k)$  by pentamers with  $h$  and  $k$  as zero or positive integers, and the third replacement identified to complete the equilateral triangle. A triangulation number  $T$  is defined as  $T=h^2+hk+k^2$  to classify the icosahedral quasi-equivalent lattices with  $T$  equal to 1, 3, 4, 7, etc. (b) the construction of a  $T=3$  icosahedron: a  $T=3$  icosahedral face is shaded in a net of seven hexamer units with  $(h=1, k=1)$  and twenty such identical faces can be arranged and folded into a three-dimensional  $T=3$  quasi-equivalent icosahedral lattice. ((a) is adapted from Johnson 1996[17] and (b) from Johnson and Speir, 1997 [15]).

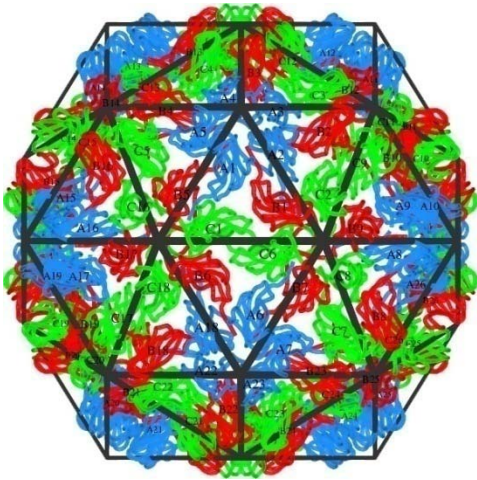
Figure I.2 shows four example icosahedral viral capsids that obey the above quasi-equivalence model, i.e.,  $T=3$  cowpea chlorotic mottle virus (CCMV), pseudo  $T=3$  poliovirus,  $T=4$  Nudaurelia capensis  $\omega$  virus (N $\omega$ V), and  $T=7$  HK97 bacteriophage. CCMV is a



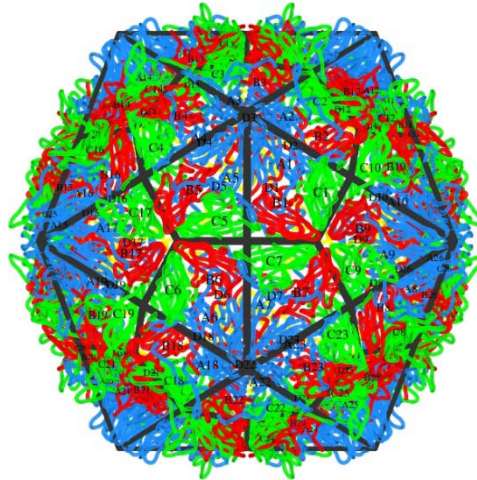
member of the bromovirus group of the Bromoviridae family and it is an extensively studied virus. Its genome consists of four positive-sense single-stranded RNA molecules, two of which are encapsulated separately in two virions and the remaining two form a third type of particle together. The crystal structure of wild-type CCMV was solved at 3.2Å resolution by X-ray crystallography (PDB ID: 1CWP) [18]. Its asymmetric unit contains three chemically identical protein subunits as colored in blue for A, red for B, and green for C as shown in Figure I.2a. Each protein subunit is composed of 190 amino acids and the viral capsid is comprised of 180 copies of the protein subunit (i.e., 60 copies of the asymmetric units) that form a 286Å diameter  $T=3$  icosahedral shell. Following from the quasi-equivalence model, CCMV capsid is formed by 12 pentamers and 20 hexamers with 5 A subunits in each pentamer, and 3 B and 3 C subunits in each hexamer. The  $T=3$  icosahedral face (i.e.,  $h=1, k=1$ ) is an equilateral triangle connecting three nearest pentamers with one hexamer located at the 3-fold axis in the middle of the triangle. Figure I.2b shows the capsid of poliovirus (mahony strain type 1, PDB ID: 2PLV) that displays a pseudo  $T=3$  icosahedral symmetry. The capsid also has three protein subunits in its asymmetric unit that are arranged similar to those in  $T=3$  CCMV capsid. But the subunit orientations in poliovirus are significantly different from those in CCMV, which cause dissimilar contacts at protein-protein interfaces and are responsible for the prominent peaks at the fivefold and threefold axes on the capsid surface [19].

The icosahedral capsid of  $T=4$  NoV (PDB ID: 1OHF) is shown in Figure I.2c. It has four protein subunits (i.e., A, B, C, and D) in its asymmetric unit, and is composed of 12 pentamers and 30 hexamers with 5 A subunits in each pentamer and 2 B, 2 C, and 2 D subunits in each

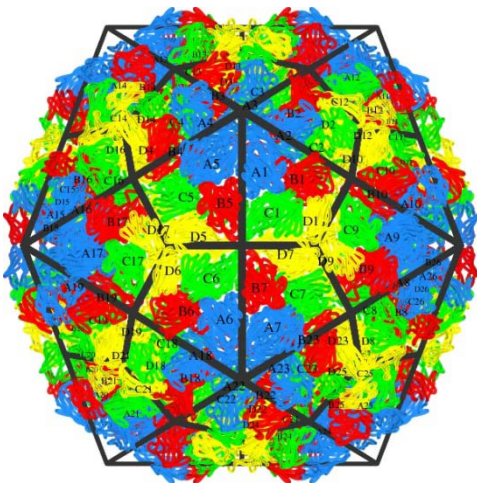
hexamer. There is one hexamer between two nearest pentamers, i.e., the capsid icosahedral face is generated by replacing two hexamers at the origin and  $(h=2, k=0)$  or  $(h=0, k=2)$  with pentamers in the net of hexamers shown in Figure I.1a. In the larger  $T=7$  HK97 bacteriophage capsid (PDB ID: 1OHG), there are seven subunits in its asymmetric subunit. The capsid is formed by 12 pentamers and 60 hexamers as shown in Figure I.2d with its icosahedral face generated by replacing the hexamer at  $(h=2, k=1)$  or  $(h=1, k=2)$  with a pentamer after one replacement at the origin.



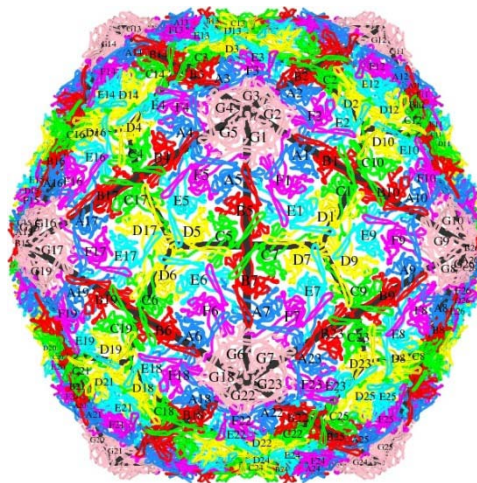
(a)  $T=3$  CCMV



(b) pseudo  $T=3$  poliovirus



(c)  $T=4$  NωV



(d)  $T=7$  HK97 bacteriophage

**Figure I.2** Four quasi-equivalent icosahedral viral capsids: (a)  $T=3$  CCMV ( $h=1, k=1$ ; PDB ID: 1CWP), (b) pseudo  $T=3$  poliovirus ( $h=1, k=1$ ; PDB ID: 2PLV), (c)  $T=4$  N $\omega$ V ( $h=2, k=0$  or  $h=0, k=2$ ; PDB ID: 1OHF), and (d)  $T=7$  HK97 bacteriophage ( $h=2, k=1$  or  $h=1, k=2$ ; PDB ID: 1OHG) (images downloaded from VIPERdb website: <http://viperdbscripps.edu/>).

While methodological problems are encountered for solving structures of viral RNA (e.g., difficulty in RNA crystallization due to its highly dynamic nature)[20], many high resolution structures (atomic or near-atomic) have been determined for viral capsids by using X-ray crystallography, cryoEM, and NMR spectroscopy [3, 4]. They serve as a starting point for investigating viral stability and dynamics, which provides insights into the self-assembly, genome packaging/releasing, structural transitions (STs) and disassembly of viruses. Their structural features make them attractive scaffolds for the study of protein-protein and protein-nucleic acid interactions [17], and the development of functional nanomaterials [21]. Their computational modeling necessities an all-atom multiscale approach as presented below. Viral simulations are of great interest for understanding their behaviors and thus the implementation of biomedical applications.

## 1.2 Viral Dynamics

The self-assembly of viruses from identical capsid proteins and genetic material is a key step in propagating their infection. The understanding the viral self-assembly, especially formation of viral capsids, is important to molecular biology as well as medical sciences and design of supramolecular synthesis. It has been proposed that delineating pathways of viral assembly could elucidate viable targets for therapeutic intervention [22]. The self-assembly

mechanism of viral capsids has been applied to synthesize functionalized supramolecules [23] and themselves have been used as molecular containers for engineered nanomaterial synthesis [21, 24-27].

The generality of the quasi-equivalence model for icosahedral viruses indicates that the origin of the icosahedral symmetry comes from the topography of the free energy landscape of viral capsids [28, 29]. Assembly studies on hepatitis B virus (HBV) showed that the free energy change for a protein-protein intersubunit contact is -3 to -4 kcal/mol, which gives -720 to -960 kcal/mol for a total of 240 contacts in the viral capsid assembly [30, 31]. However, a study has shown that the icosahedral symmetry is not a generic consequence of free energy minimization for a set of rigid capsomeres. Rather it requires optimization of internal configurations of capsid proteins [29].

In addition to the investigation of viral self-assembly thermodynamics, many experimental and theoretical studies have been undertaken to understand its chemical kinetics. For the assembly of icosahedral viral capsids, two different mechanisms have been proposed: one is the nucleation-and-growth mechanism, in which a nucleus involving protein dimers, trimers or capsomeric pentamers or hexamers is initially formed slowly, and then followed by a fast growth stage with the addition of protein subunits or dimers to form the complete capsid [32-35]; and the other is the formation mechanism similar to that for surfactant micelles, in which a relatively disordered micelle-like structure is initially formed, and through the optimization of capsid protein conformations, ordered protein capsomers emerge/reorganize to form structured icosahedral capsids [36].

It was proposed that the assembling unit of CCMV capsid is a protein dimer and the intermediate (nucleus) during capsid assembly is a  $\beta$ -hexamer of dimers (12mer) from CCMV structure study [18]. However, reexamination showed that the nucleus of CCMV capsid assembly is more likely a pentamer of dimers, rather than the hexamer of dimers, and is similar to that for brome mosaic virus (BMV) [33, 34]. For HBV, intermediates of trimer of dimers were found to be the assembly nucleus [35]. These experiments along with the theoretical studies suggest that the nucleation-and-growth mechanism is the most widely accepted assembly pathway[32]. Although many of the above efforts have been made to study viral self-assembly, its chemical kinetics still remains unclear. The intermediates during viral self-assembly need to be better characterized and their assembly pathways are underdetermined [37]. Given the many possible pathways, a fully kinetic, first-principle model is of great interest and the objective of our approach[38].

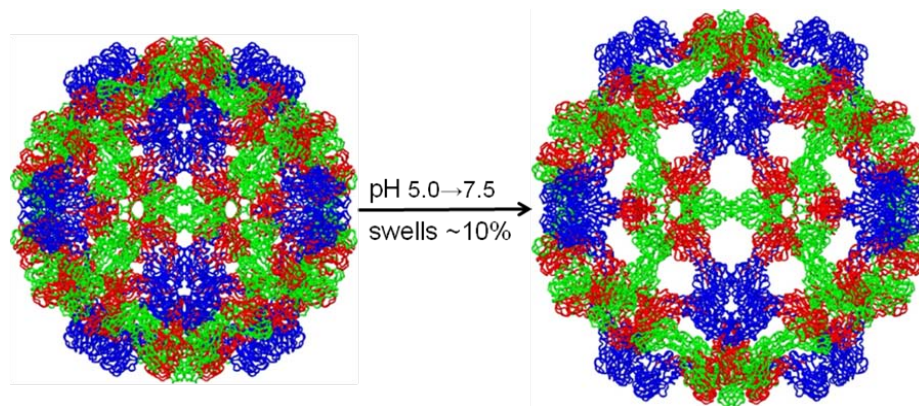
Many viruses are known to undergo STs during maturation *in vivo* or in response to changes in their microenvironment *in vitro*, such as temperature, pH, ionic strength and cation concentrations [39-44]. These are often large-scale structural changes involving reorganization of the protein subunits or capsomers in the capsid. Significant changes in the morphology, surface charge and other properties of the virus can yield metastable transient structures during viral maturation, genome releasing/packaging, or disassembly.

Native cowpea chlorotic mottle virus (CCMV) swells about 10% as pH is increased from 5.0 to 7.5 in the absence of divalent cations at low ionic strength ( $I=0.2M$ ) (Figure I.3) [1, 15, 18, 39, 40, 45], indicating the existence a free energy minimum for its swollen structure. It

disassembles at high ionic strength ( $I=1.0M$ ) [40]. In the expansion scheme proposed by Liu et al.[1], its swollen state can be generated by taking the pentamers and hexamers through the following rigid-body changes from their native configurations: translate pentamers by  $24\text{\AA}$  radially and rotate them counter-clockwise by  $9^\circ$  around their 5-fold axes; and translate hexamers by  $21\text{\AA}$  radially and rotate them counter-clockwise by  $8^\circ$  around their 3-fold axes. Irreversible poliovirus conformational changes are found in receptor-mediated cell entry, during which its coat protein VP4 and the N terminus of VP1 are externalized. Two putative cell entry intermediates (135S and 80S particles) are formed and they are both about 4% larger than the native virions (160S particles) [7, 8, 43, 44, 46]. The viral receptor behaves as a classic transition state theory catalyst, facilitating the ST from native virions to 135S intermediate particles by lowering the activation energy for the process by 50 kcal/mol [46].

Small-angle X-ray scattering (SAXS) experiments revealed that N $\phi$ V [10, 41, 47, 48] and HK97 bacteriophage [49, 50] undergo STs during capsid maturation. N $\phi$ V undergoes large conformational changes from a procapsid form ( $480\text{\AA}$  in diameter) to a compact capsid form ( $410\text{\AA}$  in diameter) when pH is decreased from 7.6 to 5.0. The transition is believed to take less than 100 ms and is accompanied by a slow autoproteolysis (taking hours) corresponding to the cleavage of 70 kDa coat proteins to 62 kDa and 8 kDa proteins. The conformational rearrangement is initially reversible until about 15% of the cleavage events are completed, at which point the particles are locked into the capsid conformation, regardless of pH [10, 47, 48]. A further study on a cleavage-defective mutant (N570T) of N $\phi$ V showed that the transition from procapsid to capsid in the mutant is reversible, and that the reverse process is much slower,

with some capsids did not reexpand after 4 days of dialysis against pH 7.6 buffer. The reexpanded procapsids display slightly different properties than the original capsid, suggesting incomplete reversibility in the transition [41]. HK97 bacteriophage capsid expands from a metastable procapsid (prohead) to a mature icosahedral capsid (head) via three steps[51]: 1) local refolding of capsid subunit N-terminal arms, 2) global expansion of the capsid involving large-scale rearrangements of capsid subunits, and 3) additional subtle and slow structural changes contributing to cross-linking of capsid subunits that stabilize the mature capsid. While greatest structural changes take place during capsid global expansion, they are triggered by local refolding of the capsid subunit N-arms. Therefore, viral STs are typically accomplished through a series of steps involving intact translation and rotation of protein structural units (e.g. pentameric and hexameric capsomers) in the capsid.



**Figure I.3** Native CCMV swells about 10% as pH is increased from 5.0 to 7.5 in the absence of divalent cations.

While native virions have genetic material within the capsid cavity, empty viral capsids can be assembled from protein subunits *in vitro* (e.g., CCMV) and probe nanoparticles (e.g. gold semiconductor or other nanoparticles) or therapeutic material (e.g. cancer drugs, small

interfering RNA (siRNA) or genes) are introduced into viral capsids for understanding the characteristics of the inner surface and the fluctuations of the capsid and biomedical applications. Empty viral capsids have also been adopted for nanomaterial synthesis by use of their STs [21, 24, 39]. The dynamics of these systems reflects the interplay of the fluctuations of the capsid and the internal particles, the interactions between the particles, and the screening or dielectric properties of the medium within the capsid. Thus the entrapped nanoparticles may form long-lived clusters, adhere to the capsid inner surface, and induce capsid STs or disassembly.

Viruses involve STs that play a critical role in the virus life cycle, such as virus attachment to cell membranes, disassembly and release of the packaged nucleic acid from viral capsid into host cell. They are attractive scaffolds for synthesizing novel nanomaterials. They provide an excellent system for validating a new computational modeling approach and studying protein-protein and protein-RNA interactions of bionanosystems.



## **Chapter 2 Viral Computational Modeling**

### **2.1 Review of Viral Computational Models**

Theoretical approaches have been developed to simulate viral properties and dynamics, including (1) computational molecular dynamics (MD) [52-66], (2) lumped coarse-grained models[67, 68], (3) symmetry-constrained models[53], (4) normal mode analysis (NMA) [69-72], and (5) Poisson-Boltzmann (PB) approaches [73, 74], but they have suffered from one or more limitations as the following.

MD has been used to model the dynamics of viruses and their interactions with drug molecules. But most of these simulations were run on only parts of a virus, e.g., a single protomer as for Human Rhinovirus (HRV) [53, 59, 60]. These studies revealed certain physical properties of viral components, such as compressibility of a HRV protomer, which were used to conjecture the ability of drug molecules to inhibit viral infection. A hypothetical short single-stranded RNA, 5'-R(PGpGpApCpUpUpCpGpGpUpCpC)-3'), was constructed to study RNA release of CCMV[66]. In the MD simulation, only one asymmetric unit of CCMV capsid was included. Results indicate that the RNA fragment loses its secondary structure and moves into the channel along the three quasi-threefold axis of CCMV capsid by free diffusion. However, these conclusions are not based on simulations of whole-capsid, full-length RNA behaviors. There has been an MD simulation on a complete satellite tobacco mosaic virus (STMV) using NAMD[75], a high-performance parallel MD code. The simulation took 50 days on a SGI 1,024 processor Altrix system using 256 processors and 128 GB of memory, but only

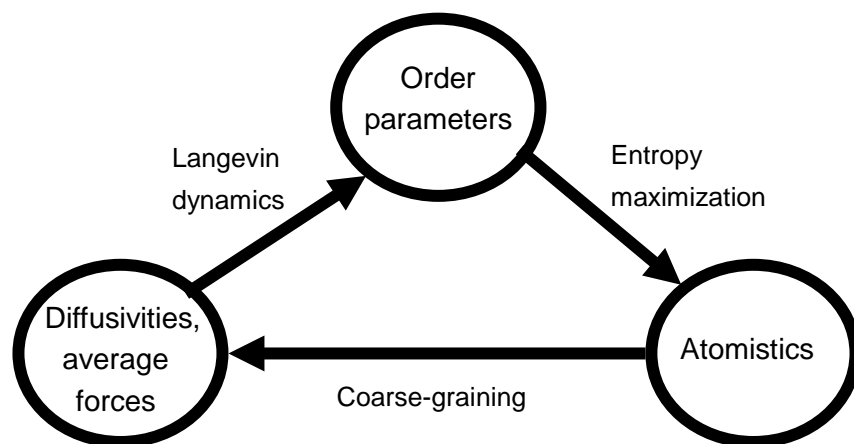
captured a physical time of 55 ns [52]. However, viral STs take milliseconds to minutes or longer to occur. With such an MD simulation, it would take about 2,500 years to obtain a physical result. Thus direct MD is not practical for whole-virus, millisecond to minute timescale modeling.

Coarse-grained models [68, 76] use a reduced description composed of lumped elements representing clusters of atoms to allow more efficient sampling and larger timestep than MD. Residue and shape based coarse-grained modeling have been used to simulate protein-lipid systems and viruses [68, 77, 78]. While they allow for long-time simulations, it (1) must be recalibrated with each new application, (2) limits its prediction power with preconception of grouping atoms into rigid clusters, (3) lacks atomistic detail as needed to account for drug molecule or cell surface receptor interaction, (4) ignores the fact that nanosystem coherent motions involve all the atoms moving collectively[79], and (5) misses the feedback between atomistic fluctuations and virus-wide structural dynamics as shown in Figure I.4.

NMA has been used to study virus capsids for insights into their structural dynamics [69-72]. Based on single-well (harmonic) potential approximations, NMA is used to explore low-frequency, large-scale structural changes. In these studies, various methods are applied to reduce the number of degrees of freedom (e.g., elastic network models and the rotation/translation of blocks method). Icosahedral symmetry was used to facilitate normal mode calculations via group theory. But an icosahedral virus does not maintain symmetry during disassembly or STs. Rather, initiation of instability is local, starting with the motion or deformation of a single structural unit, and then propagates across the virus. For pH-induced

CCMV capsid swelling, NMA provides candidate structures of intermediate and swollen CCMV, and a putative pathway for swelling. However, the method is not for simulating viral dynamics directly. NMA cannot capture the local, highly nonlinear, diffusive (friction-dominated) character of viral STs. Nor can it be readily used to study viral interactions with drugs or cell surface receptors for similar reasons.

Electrostatics of icosahedral viruses has been modeled by solving the PB equation [74, 80-82]. The electrostatic interaction between RNA and CCMV capsid has been simulated with a coarse-grained RNA model and a Monte Carlo approach [73]. In the model, each RNA nucleotide segment was treated as a sphere with a charge of  $-0.25e$  and no connectivity was enforced between the spheres. Results show that there is a very strong interaction between RNA segments and the highly positive N-terminal residues of the capsid proteins. RNA segments were predicted to form a shell close to the capsid with the highest densities associated under protein dimers. These high-density regions are connected to each other in a continuous net of triangles. Medium density of RNA is found under protein pentamers. These studies reveal viral interesting features, but they do not capture the RNA-wide structural constraints imposed by RNA connectivity. Furthermore, they have the limitations of other coarse-grained models and do not provide a general approach to whole-virus modeling. Also it is assumed that the electrical field is fixed (i.e., is imposed by the capsid), rather than being co-evolved with the RNA (and possible capsid ST) during a simulation.



**Figure I.4** Order parameters characterizing nanoscale features affect the relative probability of atomistic configurations which, in turn, mediates the forces driving order parameter dynamics.

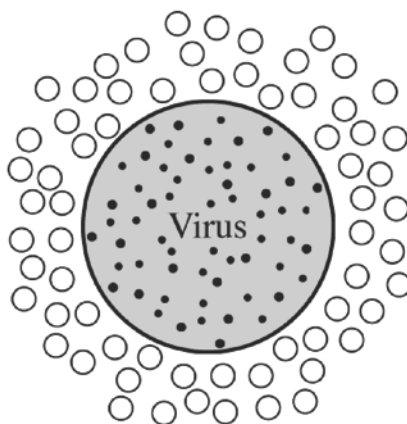
While above computational approaches reveal interesting features of viral stability and dynamics, they do not meet the challenge of whole-virus simulation for one or more of the following reasons. They (1) are limited to simulation times that are much shorter than those of biological interest (i.e., milliseconds or longer), (2) require parameter recalibration with each new application, (3) lack the atomic-scale detail needed to address viral interaction with cell receptors or drug molecules, (4) impose symmetry constraints incompatible with the local, asymmetric nature of drug or cell receptor interaction at selected sites on the virus, (5) use small amplitude vibration theory, and thereby miss the nonlinear nature of STs, (6) ignore the highly dissipative nature of nanoscale dynamics (i.e., the important role of frictional forces), and (7) miss the inter-scale feedback whereby nanoscale structural variables (or order parameters as defined in Chapter 3) affect the statistical distribution of atomic-scale fluctuations which, in turn, mediate the entropic and free energy effects that drive the dynamics of the nanoscale features (see Figure I.4). This feedback loop is central to a complete understanding of nanosystems and the true nature of their dynamics.

## 2.2 Multiscale Analysis

The central elements of a viral problem are suggested in Figure I.5, i.e. a virus is a massive and geometrically large aggregate of atoms immersed in a host medium that subjects it to frequent collisions. In the absence of macroscopic gradients in the host medium, these collisions result in a fluctuating force on the viral center-of-mass that averages out in time and across the surface of the virus but leads to a coupled Brownian translational/rotational/ST dynamics. In principle, viral dynamics can be understood in terms of a set of atoms evolving classically in an inter-atomic force field. The benefit of such an approach is that it enables one to develop a general virus model that does not need parameter calibration for each new application. The challenge is to implement this approach as a practical computational algorithm for systems with millions of atoms. A central objective is to capture atomic-scale detail and project the results into whole-virus scale responses.

Since a virus is a collection of  $N$  atoms that exchange momentum and energy with the host medium mediated by the interatomic force field, the development of projection operators [83] is not directly applicable. Because it involves memory kernels that are not readily evaluated using standard MD. These kernels can be approximated when the order parameters describing overall system behaviors evolve much slower than atomistic collisions/vibrations. However, when there is a clear timescale separation, one can directly derive stochastic equations for the order parameters, all factors within which can be evaluated by using standard MD codes. Therefore projection operators are not particularly useful. Viruses along with other large biological complexes (e.g., hemoglobin, ribosomes, cytochromes) involve processes taking place across

widely separated time and length scales. While various methodologies, such as smoothing potential energy functions[84], identification of reaction paths[85-88], simulating rare but fast events [89, 90] and “milestoning” diffusive processes [91], have been developed to accelerate the dynamics simulations, great challenge remains to capture their dynamics over long time (e.g., milliseconds to minutes) [5, 92, 93]. In the following, we adopt a multiscale approach to study viral dynamics.



**Figure I.5** The essence physical picture of a virus: a massive and geometrically large aggregate of atoms immersed in a host medium that subjects it to frequent collisions.

Multiscale analysis has been of interest since work on Brownian motion by Einstein [94-101]. It was shown that the wandering of a nanoparticle arises due to the interplay of short-scale random collisions and the large-scale, slow motions of the whole nanoparticle created by the separation in the magnitude of the mass of an atom versus that of the nanoparticle. In earlier studies, Fokker-Plank (FK) and Smoluchowski equations are derived from the Liouville equation for nanoparticles without internal atomistic structure [94-101]. Deutch and Oppenheim[95] presented an approach to study structureless Brownian particle (BP) dynamics based on the use of projection operators and a perturbation scheme developed in

the host particle/BP mass ratio. This approach set the stage for a series of studies of BP dynamics based on the Liouville equation and BP models that did not account for internal molecular structure (i.e. were structureless). Shea and Oppenheim[97] derived FP and equivalent Langevin equations for a single structureless BP in a bath of small particles via projection operators, a perturbation expansion in the mass ratio, and the assumption that gradients in the host medium are small. Peters [100, 101] derived FP equations for the coupling of rotational and translational motions of a structureless nonspherical BP near a surface. Shea and Oppenheim[98] analyzed the case of multiple BPs, introducing a number of smallness parameters including the mass ratio. Ortoleva[99] presented an approach based on a formal multiple space-time scaling approach integrated with a statistical argument derived from the BP/host particle size ratio and BP geometry that allows for a united asymptotic expansion for solving the Liouville equation; the result is an FP equation for single and multiple BPs and intra-BP dynamics. It also was shown formally that this approach leads to a set of coupled FP equations, one for each slow host mode when the slow hydrodynamics of the host medium is accounted for.

While an all-atom multiscale analysis (AMA) to structured nanoparticles seems natural, the internal dynamics of the atoms constituting a nanoparticle introduces conceptual and technical difficulties associated with a description involving both the atomistic and nanometer scale properties of these systems and the potential overcounting of degrees of freedom. To overcome these difficulties we introduce a “nanocanonical ensemble” method to facilitate the multiscale analysis of the all-atom Liouville equation. Our approach overcomes technical difficulties

associated with the removal of secular behavior, which leads to FP type equations. Our approach ensures removal of all secular behavior in the  $N$ -atom probability density and not just that of a reduced distribution. Being based on a calibrated interatomic force field, our method has the potential to yield parameter-free universal models for nanoparticle dynamics. The objective of our research is to place above viral phenomena in the framework of rigorous statistical mechanics, i.e. to show how they emerge from an all-atom description via a multiscale analysis of the Liouville equation. We seek to understand how the coherent responses of a nanoscale life form emerge from atomistic chaos. To do so we introduced novel methods for constructing collective modes with slowly-varying order parameters to capture viral coherent motions and for deriving stochastic (FP or Smoluchowski) equations of the order parameters.

We have advanced the multiscale approach to develop an AMA theory for dynamical nanoparticles [11] by (1) accounting for atomic-scale fluctuations of intraparticle structure; (2) introducing general, automatically-generated intraparticle collective modes and order parameters; (3) constructing ensembles of atomic configurations constrained to fixed values of the slowly varying order parameters, as needed to construct thermal-average forces and diffusion factors for the FP equations; (4) using a rigorous conservation law for the OP probability density, and expansion in length or time scale ratios, and the Gibbs hypothesis; and (5) establishing a criterion for the completeness of the set of order parameters. The AMA approach captures the cross-talk between the order parameters and atomistic variables rigorously as suggested in Figure I.4 [9, 11, 38, 81, 102, 103], and provides the conceptual



framework for our whole-virus simulation approach.

A multiscale coupling (MSC) approach bridged the nano and mesoscopic scales by embedding a nanoscale-level system into a mesoscopic continuum [77, 104-106]. In the case of membrane simulations, MD was used to calculate material properties to parameterize an elastic membrane model; the surface tension of a local region is then calculated and applied back to the corresponding membrane zone as a boundary condition for its MD simulation [104, 105]. This MSC approach describes the interface between a local region of a mesoscopic system and the remainder of the system. However, it does not capture the cross-talk between the rapid atomistic fluctuations and coherent nanoscale modes during system structural changes, the nonlinearity of the membrane, and the highly dissipative nature of the system. In contrast, these facets are addressed in our all-atom multiscale approach which accounts for the feedback loop of Figure I.4.

### **2.3 All-atom Multiscale Computational Modeling**

Based on concepts following from our AMA theory of nanosystem dynamics, we developed a multiscale MD/Order Parameter eXtrapolation (MD/OPX) approach for simulating viruses and other bionanosystems [12-14]. In the implementation, collective modes and order parameters constructed with basis functions (e.g., orthogonal polynomials or harmonic functions) of atomic coordinates of a reference configuration or atomic displacements between two known configurations are introduced to capture the slowly-varying nanoscale coherent motions of the system. Replica short MD runs with random atomic velocity initializations are

implemented to estimate the ensemble average rate of change in order parameters, which is then used to extrapolate the state of the system over a time period that is much longer than the  $10^{-14}$ -second timescale of fast atomic vibrations and collisions. The approach is essentially equivalent to solving the Langevin equations for stochastic dynamics of the order parameters. Since the timescale for the evolution of order parameters is much larger than that of atomic vibrations and collisions, the OP extrapolation time can be many orders of magnitude greater than the MD simulation timestep. The resulting MD/OPX algorithm accelerates MD for long-time simulation of large bionanosystems and it addresses rapid atomistic fluctuations and slowly-varying nanoscale features simultaneously underlying viral structural dynamics.

The equation free multiscale analysis (EFM) approach developed by Kevrekidis *et al.*[107] shares much of the flavor of MD/OPX, i.e., short bursts of MD simulations can be used to extrapolate coarse variables (i.e., order parameters here) over large time intervals and thus project the system over long time. It has been applied to an alanine dipeptide for molecular dynamics study by using the dipeptide dihedral angle ( $\text{N-C}_\alpha\text{-C-N}$ ) as its coarse variable[108]. However, the dipeptide is a very small system, and its dihedral angle does not appear to be a slow variable and therefore cannot be extrapolated over long time. PCA has been used to reduce the dimensionality of MD trajectories for analyzing large-scale structural changes, but PCA modes calculated for consecutive time windows of MD trajectories have been shown to display small similarity [79, 109-111]. This proves that they fail to capture system coherent motions over long time. It is thus not reliable to use PCA modes obtained from short MD runs to facilitate the long-time simulation of nanostructures because of limited sampling of long-range

correlations and forced orthogonalization of PCA modes[112-114]. In contrast, collective modes and order parameters automatically constructed with basis functions of atomic positions of a reference configuration or atomic displacements between two known configurations capture slow, nanoscale dynamics of viruses and other bionanosystems, and serve as a starting point for an all-atom multiscale analysis that justifies MD/OPX[79]. While our AMA approach and MD/OPX with their underlying all-atom description of nanosystems enable the use of a universal inter-atomic force field and can be applied to all dynamical nanoparticles [11]. Here, we are focused on understanding the dynamics of icosahedral viruses, such as CCMV, N $\omega$ V and poliovirus. Methods including AMA formulation of the  $N$ -atom Liouville equation with construction of order parameters capturing viral coherent motions, derivation of the stochastic equations for coarse-grained dynamics of order parameters, a multiscale single-order parameter model for viral dilatational STs, and MD/OPX algorithm and implementation are presented in Part II. In Part III, AMA and MD/OPX are applied to study the dynamics of icosahedral viruses, notably N $\omega$ V and CCMV, with their ST pathway and mechanisms revealed and discussed. Conclusions and the usage of multiscale simulation predictions to guide experiments are provided in Part IV.

## Part II. Methods

### Overview

All-atom multiscale theoretical and computational methods for studying viral dynamics are presented in this part. In Chapter 3, all-atom multiscale analysis (AMA) is formulated for the  $N$ -atom Liouville equation of dynamical nanoparticles with viruses as our special interest[11]. Slow variables (i.e., order parameters) are identified to capture the internal structure of nanoparticles. By removing all secular behavior of the  $N$ -atom probability density using a “nanocanonical ensemble” approach, AMA of the Liouville equation for structured nanoparticles leads to a stochastic (Fokker-Planck (FP) or Smoluchowski) equation yielding the coarse-grained evolution of the order parameters. Systematic construction of collective modes and order parameters from atomic variables for capturing nanosystem coherent motions are provided [79] and generalizations of the AMA approach are also presented. In Chapter 4, order parameters capturing viral dilatational structural transitions are constructed, a FP equation for their long timescale dynamics is derived by following the AMA formulation in the inertial limit, and a transition state ansatz is presented for obtaining approximate solutions to the FP equation. In Chapter 5, the multiscale Molecular Dynamics/Order Parameter eXtrapolation (MD/OPX) approach is presented along with its implementation details [12-14].

## Chapter 3 All-atom Multiscale Analysis (AMA)

### 3.1 AMA of the $N$ -atom Liouville Equation

The scaling approach and challenges involved in all-atom multiscale analysis (AMA) of the Liouville equation for structured nanoparticles (e.g., icosahedral viruses) are now introduced. Consider an  $N$ -atom system comprised of a nanoparticle of  $N^*$  atoms and a host medium of  $N - N^*$  atoms. To characterize the differences in length, mass and time scales involved in a nanosystem we introduce a unifying smallness parameter  $\varepsilon$  ( $\ll 1$ ). Thus the nanoparticle center-of-mass (COM) momentum and position are  $\varepsilon^{-1}\bar{P}$  and  $\varepsilon^{-1}\bar{R}$ . With this, the  $N$ -atom probability density  $\rho$  is considered to be a function of the scaled COM momentum and position, as well as the momenta and positions  $\Gamma = \{\bar{p}_i, \bar{r}_i, i = 1, \dots, N\}$  of the  $N$  atoms in the system. A key point is that the multiple distinct dependencies in  $\rho$  does not imply a violation of the restriction on the number of degrees of freedom; rather this is to state that  $\rho$  has multiple scale character – here the long range migration of the large mass nanoparticle in the presence of the rapidly fluctuating atomic variables  $\Gamma$ .

Using the chain rule,  $\rho$  in the above multiscale form satisfies (see Appendix A)

$$\frac{\partial \rho}{\partial t} = (\mathcal{L}_0 + \varepsilon^2 \mathcal{L}_1) \rho \quad (3.1.1)$$

$$\mathcal{L}_0 = - \sum_{i=1}^N \left[ \frac{\bar{p}_i}{m_i} \cdot \frac{\partial}{\partial \bar{r}_i} + \bar{F}_i \cdot \frac{\partial}{\partial \bar{p}_i} \right] \quad (3.1.2)$$

$$\mathcal{L}_1 = - \frac{\bar{P}}{m} \cdot \frac{\partial}{\partial \bar{R}} - \bar{f} \cdot \frac{\partial}{\partial \bar{P}}. \quad (3.1.3)$$

In this formulation we have introduced the dimensionless parameter  $\varepsilon$  such the  $\varepsilon^2 = m/m^*$  for typical atomic mass  $m$  and nanoparticle mass  $m^*$ . The net force  $\varepsilon \bar{f}$  on the nanoparticle can be written in terms of the individual atomic forces  $\bar{F}_i$ :

$$\varepsilon \bar{f} = \sum_{i=1}^N \Theta_i \bar{F}_i \quad (3.1.4)$$

where  $\Theta_i$  is one for atoms within the nanoparticle and zero for host atoms. The net force on the nanoparticle is taken to be small; for the quasi-equilibrium states of interest there is much cancellation of the forces on the nanoparticle from those on individual atoms. It is further assumed that external forces are weak. Alternative scalings for nanoparticle mass, momentum, position and force can also be adopted to investigate other classes of behavior, but the above will serve to illustrate the present methodology. That derivatives in  $\mathcal{L}_0$  are at constant  $\bar{P}$  and  $\bar{R}$ , while those in  $\mathcal{L}_I$  are at constant  $\Gamma$ , does not imply that  $\Gamma$ ,  $\bar{P}$  and  $\bar{R}$  are independent degrees of freedom; rather they arise due to our attempt to express  $\rho$ 's multiscale character. Thus any integration over all states of the system would be over the dependence on  $\Gamma$ , including that in  $\bar{P}$  and  $\bar{R}$ .

The multiscale development proceeds by expanding  $\rho$  in a series in  $\varepsilon^2$ :

$$\rho = \sum_{n=0}^{\infty} \rho_n(\Gamma, \bar{P}, \bar{R}, t_0, \underline{t}) \varepsilon^{2n} \quad (3.1.5)$$

where  $t_n = \varepsilon^{2n} t$  and  $\underline{t} = \{t_1, t_2, \dots\}$  is the collection of times characterizing the slow behaviors, i.e. that evolve on times of order  $\varepsilon^{-2}, \varepsilon^{-4}$  and longer.

The lowest order distribution is assumed to reflect the quasi-equilibrium nature of the

biological phenomena of interest, i.e.  $\rho_0$  is independent of  $t_0$ . Using the chain rule for  $\partial/\partial t$  (so that  $\partial\rho/\partial t = \partial\rho/\partial t_0 + \varepsilon^2\partial\rho/\partial t_1 + \dots$ ), the lowest order problem becomes

$$\mathcal{L}_0\rho_0 = 0. \quad (3.1.6)$$

This equation yields solutions which are functions of the Hamiltonian  $H_0$  that generates  $\mathcal{L}_0$ , i.e.

$$H_0 = \sum_{i=1}^N \frac{p_i^2}{2m_i} + V \quad (3.1.7)$$

where  $V$  is the  $N$ -atom potential. For example,  $\rho_0$  could be the canonical distribution  $\exp(-\beta H_0)$

which is normalized by the partition function  $Q$ :

$$Q = \int^* d^{6N} \Gamma \exp(-\beta H_0) \quad (3.1.8)$$

where  $d^{6N} \Gamma = d^3 p_1 d^3 r_1 \dots d^3 p_N d^3 r_N$ . The  $*$  on the integration implies that  $\bar{R}$  and  $\bar{P}$  must be fixed, i.e. for any quantity  $A$ ,

$$\int^* d^{6N} \Gamma A = \int d^{6N} \Gamma \delta\left(\bar{P} - \varepsilon \sum_{i=1}^N \Theta_i \bar{p}_i\right) \delta\left(\bar{R} - \varepsilon \sum_{i=1}^N \Theta_i m_i \bar{r}_i / m^*\right) A. \quad (3.1.9)$$

Such integrals can be accomplished via Monte Carlo methods and a constant  $\bar{P}, \bar{R}$ -constrained generation of  $\Gamma$  configurations. The  $O(\varepsilon^0)$  analysis typically concludes with the introduction of the slowly varying factor  $W(\bar{P}, \bar{R}, \underline{t})$ , i.e.  $\rho_0 = Q^{-1} \exp(-\beta H_0) W$ . In Sects. 3.2 and 3.3 we demonstrate how the above formulation can be placed on more rigorous footing.

A difficulty with the above formulation is that while solving (3.1.6) one must keep the delta functions in (3.1.9) and the constrained integrations as part of the higher order multiscale analysis.

For example, a key difficulty arises in analyzing the  $O(\varepsilon^2)$  equation:

$$\left( \frac{\partial}{\partial t_0} - \mathcal{L}_0 \right) \rho_1 = -\frac{\partial \rho_0}{\partial t_1} + \mathcal{L}_1 \rho_0. \quad (3.1.10)$$

To solve this equation one must recognize that the RHS can have contributions from the null space of  $\mathcal{L}_0$ . In this recognition one would like to arrive at an expression for  $\partial W / \partial t_1$  by ensuring that the resulting secular behavior (i.e. divergence at large  $t_0$ ) is removed. The difficulty is that in the classic approach an expression for  $\partial W / \partial t_1$  follows from applying  $\int^* d^{6N} \Gamma$  to both sides of (3.1.10) and then using the fact that acceptable probability distributions vanish as  $|\bar{p}_i| \rightarrow \infty$ , and invoking periodic boundary conditions on the  $\bar{r}_i$ -dependence of the probability. In studies of structureless nanoparticles this directly implies  $\int d^{6N} \Gamma \mathcal{L}_0 A = 0$  for any distribution  $A$  and wherein particle  $N+1$  is the structureless nanoparticle. This yields the Fokker-Planck equation when computations are carried out to  $O(\varepsilon^4)$ . The strategy is not clear for the structured nanoparticle wherein the constrained integration  $\int^* d^{6N} \Gamma$  must be used to preserve the  $\bar{P}, \bar{R}$ -dependence of  $\rho$ . This and related difficulties stem from the fact that unlike for the traditional approach, the structured nanoparticle's atomic coordinates are already a complete set of dynamical variables so that either  $\bar{P}, \bar{R}$  are redundant or ignoring their relation to  $\Gamma$  would allow for atomic configurations that are inconsistent with  $\bar{P}$  and  $\bar{R}$ . Hence to avoid overcounting the number of degrees of freedom the  $\int^* d^{6N} \Gamma$  constrained integrals must be used. But then the boundary conditions on the  $\bar{p}_i, \bar{r}_i$  cannot be readily utilized.



A more fundamental question also arises. Even if one can derive an expression for  $\partial W/\partial t_l$  via a  $\Gamma$ -integration, the question remains that although secular behavior is removed from a reduced aspect of  $\rho_l$  (e.g.  $\int^* d^{6N} \Gamma \rho_l$ ), it is still not demonstrated that all the secular behavior in  $\rho_l$  itself has been removed. What is needed is an approach wherein the nanoparticle COM degrees of freedom, or something related to them, are not independent dynamical variables which must ultimately be related to  $\Gamma$ , and a method that will remove all the secular behavior without resorting to  $\Gamma$  integration. Such an approach is developed in Sects. 3.2 and 3.3.

The  $N$ -atom potential  $V(\bar{r}_l, \dots, \bar{r}_N)$  does not explicitly depend on nanoparticle COM position  $\bar{R}$ . This is in contrast with the case of a structureless nanoparticle which can be taken as “atom”  $N+1$  whereby  $\bar{r}_{N+1}$  appears in  $V$  explicitly. To resolve this issue, introduce a set of modified coordinates  $\bar{s}_i$  such that  $\bar{s}_i = \bar{r}_i - \varepsilon^{-1} \bar{R} \Theta_i$ . Thus the atoms in the nanoparticle are described in the relative frame while those of the host medium are in the laboratory frame. The  $\bar{s}_i$  are functions of  $\bar{r}_l, \dots, \bar{r}_N$ ,  $\bar{s}_i = \bar{s}_i(\bar{r}_l, \dots, \bar{r}_N)$  and, like  $\bar{R}$ , are dependent dynamical variables. With this, the potential  $V$  can be expressed in the form

$$V\left(\left\{\varepsilon^{-1} \bar{R} \Theta_i + \bar{s}_i, i = l, \dots, N\right\}\right) \equiv U\left(\bar{r}_l, \dots, \bar{r}_N, \bar{R}\right) \quad (3.1.11)$$

so that the explicit dependence of  $U$  on the  $\bar{r}_i$  is only through relative configuration  $\{\bar{s}_l, \dots, \bar{s}_N\}$ .

With this the particle derivative of  $V$  with respect to  $\bar{R}$  at constant nanoparticle internal configuration is given by

$$\left(\frac{\partial V}{\partial \bar{R}}\right)_{\bar{s}_l, \dots, \bar{s}_N} = -\frac{1}{\varepsilon} \sum_{i=l}^N \Theta_i \bar{F}_i = -\bar{f}. \quad (3.1.12)$$

Finally note that the  $\bar{s}_i$  are not independent, i.e.  $m_I \bar{s}_I \Theta_I + \dots m_N \bar{s}_N \Theta_N = \bar{0}$ . This result forwards the multiscale analysis of the following sections.

### 3.2 Nanocanonical Ensemble Approach

In Sect. 3.1 we considered the nanoparticle to have a particular momentum and position that characterized the lowest order solution to the Liouville equation. However, it would seem more consistent, and is found to have technical advantages, to cast the problem in terms of the dynamics of the ensemble of uncertain nanoparticle states – i.e. to introduce a probability distribution with parameters that characterize the statistics of an ensemble of systems in various nanoparticle states. In the process of developing this notion, we show that the conceptual and technical difficulties encountered in Sect. 3.1 are overcome.

To solve the Liouville equation via a multiscale approach, we first seek a quasi-equilibrium solution to the lowest order equation:

$$\mathcal{L}_0 \rho_0 = 0. \quad (3.2.1)$$

As  $\mathcal{L}_0 \bar{P}, \mathcal{L}_0 \bar{R}$  and  $\mathcal{L}_0 H_0$  vanish, we seek a solution of (3.2.1) that is a function of  $H_0$ ,  $\bar{P}$  and  $\bar{R}$ . The lowest order solution of interest here is constrained such that the ensemble average of the  $\Gamma$ -dependent quantities  $H_0$ ,  $\bar{R}$  and  $\bar{P}$  have prescribed values (that may vary on the slow timescale), and hence  $H_0$ ,  $\bar{P}$  and  $\bar{R}$  themselves are not fixed. This difference from the approach of Sect. 3.1 yields great technical and conceptual advantages.

Using information theory, one obtains the following distribution through the imposition of the

aforementioned averages:

$$\hat{\rho}_{\bar{\mu}, \bar{\kappa}} = \frac{e^{-\beta H_0 + \bar{\mu} \cdot \bar{P} + \bar{\kappa} \cdot \bar{R}}}{\Xi(\beta, \bar{\mu}, \bar{\kappa})}. \quad (3.2.2)$$

This lowest order solution depends on the dynamical variables  $\Gamma$  through  $H_0$ ,  $\bar{P}$  and  $\bar{R}$ . It is implied by entropy maximization through the introduction of the Lagrange multipliers  $\beta$ ,  $\bar{\mu}$  and  $\bar{\kappa}$ . We term  $\bar{\mu}$  and  $\bar{\kappa}$  reciprocal quantities in that they will be found to play roles similar to Fourier transform variables, e.g.  $\bar{\kappa}$  will, through inverse transformation, imply a derivative with respect to COM position in a Fokker-Planck equation, and similarly for the reciprocal momentum  $\bar{\mu}$ . As is familiar in classical systems, the quasi-equilibrium distribution  $\hat{\rho}_{\bar{\mu}, \bar{\kappa}}$  separates into a momentum and a position factor, and similarly for the partition function  $\Xi(\beta, \bar{\mu}, \bar{\kappa})$ .

The partition function  $\Xi(\beta, \bar{\mu}, \bar{\kappa})$  in (3.2.2) is given by

$$\Xi(\beta, \bar{\mu}, \bar{\kappa}) = \int d^{6N} \Gamma \exp[-\beta H_0 + \bar{\mu} \cdot \bar{P} + \bar{\kappa} \cdot \bar{R}]. \quad (3.2.3)$$

The integration over all atomic degrees of freedom is unrestricted – i.e. one is not burdened with the complexity of restricting the coordinates of the atoms in the nanoparticle to a fixed COM location, and similarly for the momenta. In expressions like (3.2.3)  $\bar{P}$  and  $\bar{R}$  depend on all the viral  $\bar{r}_i, \bar{p}_i$ . This nanocanonical ensemble ensures that atomic configurations which correspond to large departures of the COM coordinates from the specified  $\bar{\mu}, \bar{\kappa}$ -dependent average values will be improbable.

Let  $\langle \dots \rangle_{\bar{\mu}, \bar{\kappa}}$  imply a  $\hat{\rho}_{\bar{\mu}, \bar{\kappa}}$ -weighted average over all configurations  $\Gamma$  :

$\langle A \rangle_{\bar{\mu}, \bar{\kappa}} = \int d^{6N} \Gamma A \hat{\rho}_{\bar{\mu}, \bar{\kappa}}$  for any  $\Gamma$ -dependent quantity  $A$ . The averages  $\langle \bar{P} \rangle_{\bar{\mu}, \bar{\kappa}}$  and  $\langle \bar{R} \rangle_{\bar{\mu}, \bar{\kappa}}$  are given by

$$\langle \bar{P} \rangle_{\bar{\mu}, \bar{\kappa}} = \frac{\partial \ln \Xi}{\partial \bar{\mu}}, \langle \bar{R} \rangle_{\bar{\mu}, \bar{\kappa}} = \frac{\partial \ln \Xi}{\partial \bar{\kappa}}. \quad (3.2.4)$$

The lowest order equation (3.2.1) is linear; thus it admits the more general supposition solution

$$\rho_0 = \int d^3 \mu d^3 \kappa \hat{\rho}_{\bar{\mu}, \bar{\kappa}} \Psi(\bar{\mu}, \bar{\kappa}, \underline{t}), \quad (3.2.5)$$

for reciprocal distribution  $\Psi$ . As  $\rho_0$  is unit normalized, and  $\hat{\rho}_{\bar{\mu}, \bar{\kappa}}$  is normalized by construction, then

$$\int d^3 \mu d^3 \kappa \Psi(\bar{\mu}, \bar{\kappa}, \underline{t}) = 1. \quad (3.2.6)$$

With this, ensemble averages indicated by superscript  $m$  can be obtained. For example

$$\bar{P}^m = \int d^3 \mu d^3 \kappa \langle \bar{P} \rangle_{\bar{\mu}, \bar{\kappa}} \Psi(\bar{\mu}, \bar{\kappa}, \underline{t}) \quad (3.2.7)$$

$$\bar{f}^m = \int d^3 \mu d^3 \kappa \int d^{6N} \Gamma \bar{f} \hat{\rho}_{\bar{\mu}, \bar{\kappa}} \Psi(\bar{\mu}, \bar{\kappa}, \underline{t}) = \int d^3 \mu d^3 \kappa \langle \bar{f} \rangle_{\bar{\mu}, \bar{\kappa}} \Psi(\bar{\mu}, \bar{\kappa}, \underline{t}). \quad (3.2.8)$$

One may relate  $\Psi(\bar{\mu}, \bar{\kappa}, \underline{t})$  and the nanoparticle momentum-position density  $W$ . By definition

$$W(\bar{p}, \bar{r}, \underline{t}) = \int d^3 \mu d^3 \kappa \langle \delta(\bar{p} - \bar{P}) \delta(\bar{r} - \bar{R}) \rangle_{\bar{\mu}, \bar{\kappa}} \Psi(\bar{\mu}, \bar{\kappa}, \underline{t}). \quad (3.2.9)$$

The definition of the delta function implies

$$W(\bar{p}, \bar{r}, \underline{t}) = \int d^{6N} \Gamma \int d^3 \mu d^3 \kappa \delta(\bar{p} - \bar{P}) \delta(\bar{r} - \bar{R}) \frac{\exp(-\beta H_0 + \bar{\mu} \cdot \bar{p} + \bar{\kappa} \cdot \bar{r})}{\Xi(\beta, \bar{\mu}, \bar{\kappa})} \Psi(\bar{\mu}, \bar{\kappa}, \underline{t}). \quad (3.2.10)$$

Defining  $Q(\beta, \bar{p}, \bar{r})$  via

$$Q(\beta, \bar{p}, \bar{r}) = \int d^{6N} \Gamma \delta(\bar{p} - \bar{P}) \delta(\bar{r} - \bar{R}) \exp(-\beta H_0), \quad (3.2.11)$$

one obtains

$$\frac{W(\bar{p}, \bar{r}, \underline{t})}{Q(\beta, \bar{p}, \bar{r})} = \int d^3 \mu d^3 \kappa e^{\bar{\mu} \cdot \bar{p} + \bar{\kappa} \cdot \bar{r}} \frac{\Psi(\bar{\mu}, \bar{\kappa}, \underline{t})}{\Xi(\beta, \bar{\mu}, \bar{\kappa})} \quad (3.2.12)$$

so that  $W/Q$  and  $\Psi/\Xi$  are related via a bilateral Laplace transform. The inverse of this relationship casts  $\Psi/\Xi$  in terms of an interesting integral of  $W/Q$  over imaginary nanoparticle momenta and positions (see Appendix B).

The results (3.2.12) and (3.2.5) imply

$$\rho_0 = \frac{e^{-\beta H_0}}{Q(\beta, \bar{P}, \bar{R})} W(\bar{P}, \bar{R}, t). \quad (3.2.13)$$

Note that here  $W$  depends on  $\bar{P}$  and  $\bar{R}$ , which, in turn, are functions of  $\Gamma$  and hence are dependent dynamical variables. In contrast,  $\bar{p}$  and  $\bar{r}$  in (3.2.10) are not dynamical variables (e.g.  $\partial \bar{p} / \partial \bar{p}_i$  is zero).

As we are interested in developing an equation of motion for  $W$ , we adopt (3.2.13) as the starting point of our analysis rather than the  $\Psi$  formulation. This is shown to rather directly lead to an equation for  $W$  in the next section wherein we show that this follows by removing secular behavior using a method that does not involve an integration over  $\Gamma$ .

Before proceeding further, we examine the structure of  $Q$  in more detail. First we introduce a configurational factor  $Z(\beta, \bar{R})$  such that

$$Z(\beta, \bar{R}) = \int d^3 r_1 \cdots d^3 r_N \delta\left(\bar{R} - \varepsilon \sum_{i=1}^N m_i \bar{r}_i \Theta_i / m^*\right) e^{-\beta V}, \quad (3.2.14)$$

where the  $\bar{R}$ -dependence of  $V$  is noted in (3.1.11). Similarly we introduce a momentum factor  $Y$  such that

$$Y(\beta, \bar{P}) = \int d^3 p_1 \cdots d^3 p_N \delta \left( \bar{P} - \varepsilon \sum_{i=1}^N \Theta_i \bar{p}_i \right) e^{-\beta K}, \quad (3.2.15)$$

$$K = \sum_{i=1}^N p_i^2 / 2m_i. \quad (3.2.16)$$

With this,  $Q(\beta, \bar{P}, \bar{R}) = Y(\beta, \bar{P}) Z(\beta, \bar{R})$ . Introducing the new variables  $\bar{\pi}_i$  via  $\bar{p}_i = \bar{\pi}_i + \varepsilon m_i \Theta_i \bar{P} / m$ , and neglecting terms that vanish as  $\varepsilon \rightarrow 0$ , one finds  $Y = \int d^3 \pi_1 \cdots d^3 \pi_N \delta \left( \sum_{i=1}^N \Theta_i \bar{\pi}_i \right) \exp \left( -\beta \sum_{i=1}^N \frac{\pi_i^2}{2m_i} \right)$  so that  $Y$ , and hence  $Q$ , is independent of  $\bar{P}$ . Thus we write  $Q(\beta, \bar{R})$  in the developments of the next section. From its definition and (3.1.12)

$$\partial \ln Q / \partial \bar{R} = \beta f^{Th} \quad (3.2.17)$$

where  $f^{Th}$  is the scaled net force on the nanoparticle averaged over all atomic configurations consistent with the nanoparticle having COM at  $\varepsilon^{-1} \bar{R}$ .

### 3.3 Derivation of a Fokker-Planck equation for Stochastic Dynamics of Order Parameters

The objective of the present section is to complete the nanocanonical derivation of a Fokker-Planck equation for structured nanoparticles. The starting point is the lowest order solution (3.2.13). As pointed out at the end of the previous section,  $Q$  does not depend on  $\bar{P}$  and

thus is written  $Q(\beta, \bar{R})$  henceforth. With this we rewrite  $\rho_0$  as

$$\rho_0 = \frac{e^{-\beta H_0}}{Q(\beta, \bar{R})} W(\bar{P}, \bar{R}, \underline{t}) \equiv \hat{\rho} W. \quad (3.3.1)$$

By writing (3.3.1) in this form, we do not take  $\bar{P}$  and  $\bar{R}$  to be independent variables. The apparent independence of  $\bar{P}$  and  $\bar{R}$ , as in the partial derivatives in  $\mathcal{L}_0$  and  $\mathcal{L}_I$ , is only to highlight the distinct ways (direct in  $\Gamma$  itself and indirect through  $\bar{P}$  and  $\bar{R}$ ) that  $\rho$  depends on the all-atom state  $\Gamma$ . To be sure one could take  $\bar{P}$  and  $\bar{R}$  to be dynamical variables and then eliminate six atomic variables (e.g.  $\bar{p}_I$  and  $\bar{r}_I$ ). But this would lead to a cumbersome formulation with much tedious bookkeeping and complexity of computation.

An equation for  $W$  is obtained via an examination of higher order terms in the Liouville equation. To  $O(\varepsilon^2)$  and in light of (3.1.11 and 3.1.12) the Liouville equation implies, upon dropping the 0 on  $t_0$  henceforth,

$$\left( \frac{\partial}{\partial t} - \mathcal{L}_0 \right) \rho_I = - \left[ \frac{\partial W}{\partial t_I} + \beta \frac{\bar{P}}{m} \cdot (\bar{f} - \bar{f}^{Th}) W + \frac{\bar{P}}{m} \cdot \frac{\partial W}{\partial \bar{R}} + \bar{f} \cdot \frac{\partial W}{\partial \bar{P}} \right] \hat{\rho} \equiv G(\Gamma, \bar{P}, \bar{R}; W) \rho \quad (3.3.2)$$

where  $\bar{f}^{Th}$  arises from a  $\partial Q / \partial \bar{R}$  term as in (3.2.17): in the present context

$$\bar{f}^{Th} = \int^* d^{6N} \Gamma \hat{\rho} \bar{f}. \quad (3.3.3)$$

This yields, upon taking  $\rho_I$  to be zero at  $t = 0$ , a solution of the form

$$\rho_I = \int_0^t dt' e^{\mathcal{L}_0(t-t')} G(\Gamma, \bar{P}, \bar{R}; W) \hat{\rho}. \quad (3.3.4)$$

Assembling the above results and using  $\mathcal{L}_0 H_0$ ,  $\mathcal{L}_0 \bar{P}$  and  $\mathcal{L}_0 \bar{R}$  vanish, and

$$e^{\mathcal{L}_0(t-t')} \left[ G(\Gamma, \bar{P}, \bar{R}; W) \hat{\rho} \right] = \rho e^{\mathcal{L}_0(t-t')} G(\Gamma, \bar{P}, \bar{R}; W) \text{ yields}$$

$$\rho_I = - \left\{ t \frac{\partial W}{\partial t_I} + \int_0^t dt' e^{\mathcal{L}_0(t-t')} \left[ \frac{\bar{P}}{m} \cdot \frac{\partial W}{\partial \bar{R}} + \bar{f} \cdot \frac{\partial W}{\partial \bar{P}} + \beta \frac{\bar{P}}{m} \cdot (\bar{f} - \bar{f}^{Th}) W \right] \right\} \hat{\rho}. \quad (3.3.5)$$

Secular behavior is seen in the  $\partial W / \partial t_I$  term. However there are others as well. For example,

with the change of variables  $t' = t + \tau$  one has

$$\int_0^t dt' e^{\mathcal{L}_0(t-t')} \bar{f} = \int_{-t}^0 d\tau \bar{f}(\tau) \approx t \bar{f}^{Th} \text{ for large } t. \quad (3.3.6)$$

A fundamental hypothesis of statistical mechanics asserts the equivalence of time and ensemble averages in quasi-equilibrium systems. This implies

$$\lim_{t \rightarrow \infty} \frac{1}{t} \int_{-t}^0 d\tau \bar{f}(\tau) = \bar{f}^{Th}. \quad (3.3.7)$$

where  $\bar{f}(\tau) = e^{-\mathcal{L}_0 \tau} \bar{f}$ . Hence secular behavior is removed from  $\rho_I$  if

$$\frac{\partial W}{\partial t_I} = - \frac{\bar{P}}{m} \cdot \frac{\partial W}{\partial \bar{R}} - \bar{f}^{Th} \cdot \frac{\partial W}{\partial \bar{P}}. \quad (3.3.8)$$

Combining the above results, one obtains

$$\rho_I = - \hat{\rho} \int_0^t dt' e^{\mathcal{L}_0(t-t')} (\bar{f} - \bar{f}^{Th}) \cdot \left( \beta \frac{\bar{P}}{m} + \frac{\partial}{\partial \bar{P}} \right) W, \quad (3.3.9)$$

and  $\rho_I = 0$  as  $t \rightarrow \infty$ .

Note that these results were obtained without integrating the Liouville equation at  $O(\varepsilon^2)$  over  $\Gamma$  and using the fact that  $\rho$  vanishes as  $|\bar{p}_i| \rightarrow \infty$  and is periodic in the  $\bar{r}_i$ . This traditional method does not guarantee that the full  $\rho_I$  is free of secular behavior. The present method addresses the full time dependence of  $\rho_I$ , ensures that it, and not an integrated reduction of it, has no secular behavior. Furthermore, the technical difficulties due to the fact that the  $\Gamma$  integrals



applied must be restricted as in (3.3.3) in order to preserve the number of degrees of freedom.

To  $O(\varepsilon^4)$  one finds

$$\left(\frac{\partial}{\partial t} - \mathcal{L}_0\right)\rho_2 = -\frac{\partial\rho_0}{\partial t_2} - \frac{\partial\rho_1}{\partial t_1} + \mathcal{L}_1\rho_1. \quad (3.3.10)$$

Again with  $\rho_2 = 0$  at  $t = 0$ , we find

$$\rho_2 = \int_0^t dt' e^{\mathcal{L}_0(t-t')} \left( -\hat{\rho} \frac{\partial W}{\partial t_2} - \frac{\partial\rho_1}{\partial t_1} + \mathcal{L}_1\rho_1 \right). \quad (3.3.11)$$

Using (3.3.9) for  $\rho_1$  we obtain

$$\rho_2 = \hat{\rho} \left\{ -t \frac{\partial W}{\partial t_2} + \int_0^t dt' e^{\mathcal{L}_0(t-t')} \left( \frac{\bar{P}}{m} \cdot \frac{\partial}{\partial \bar{R}} + \bar{f} \cdot \frac{\partial}{\partial \bar{P}} \right) \int_0^{t'} dt'' e^{\mathcal{L}_0(t'-t'')} (\bar{f} - \bar{f}^{Th}) \cdot \left( \beta \frac{\bar{P}}{m} + \frac{\partial}{\partial \bar{P}} \right) W \right\} + \dots \quad (3.3.12)$$

The missing term from  $\partial\rho_1/\partial t_1$  only involves one  $\bar{f} - \bar{f}^{Th}$  term and is seen upon invoking (3.3.7) to make no secular contribution. Using arguments as for  $\rho_1$  to ensure that  $\rho_2$  has no secular behavior, and the new variables  $\tau, \sigma$  such that  $t' = t + \tau, t'' = t + \tau + \sigma$ , we obtain

$$\frac{\partial W}{\partial t_2} = \lim_{t \rightarrow \infty} \frac{1}{t} \int_{-t}^0 dt \tau e^{-\mathcal{L}_0 \tau} \left( \frac{\bar{P}}{m} \cdot \frac{\partial}{\partial \bar{R}} + \bar{f} \cdot \frac{\partial}{\partial \bar{P}} \right) \int_{-t-\tau}^0 d\sigma e^{-\mathcal{L}_0 \sigma} (\bar{f} - \bar{f}^{Th}) \cdot \left( \beta \frac{\bar{P}}{m} + \frac{\partial}{\partial \bar{P}} \right) W. \quad (\text{IV}13)$$

As  $\mathcal{L}_0 \hat{\rho} = 0$  implies that only the  $\bar{f}$  factors are affected by the evolution operators has been used in obtaining this result. To complete the analysis, consider the key term

$$t J_{\alpha_1 \alpha_2} \equiv \int_{-t}^0 d\tau e^{-\mathcal{L}_0 \tau} f_{\alpha_1} \int_{-t-\tau}^0 d\sigma (f_{\alpha_2}(\sigma) - f_{\alpha_2}^{Th}), \quad (3.3.14)$$

where  $f_{\alpha_2}(\sigma)$  is the  $\alpha_2$  Cartesian component of  $\bar{f}$  evolved to time  $\sigma$ . For large but finite  $t$  the first integral with the evolution operator  $\exp(-\mathcal{L}_0 \tau)$  is approximately  $t$  times the ensemble average, i.e.

$$tJ_{\alpha_1\alpha_2} \approx t \int_{-t-\tau}^0 d\sigma \left( f_{\alpha_1}(0) \left( f_{\alpha_2}(\sigma) - f_{\alpha_2}^{Th} \right) \right)^{Th}. \quad (3.3.15)$$

Since  $\tau$  is fixed but  $t$  is arbitrarily large, and the integrand vanishes as  $\sigma$  gets large and negative, (i.e.  $\bar{f}$  and  $\bar{f}(\sigma)$  are independent random variables for large  $\sigma$ ), we obtain

$$J_{\alpha_1\alpha_2} \approx \int_{-\infty}^0 d\sigma \left( f_{\alpha_1}(0) \left[ f_{\alpha_2}(\sigma) - f_{\alpha_2}^{Th} \right] \right)^{Th}. \quad (3.3.16)$$

Using this result, secular behavior in  $\rho_2$  is removed when

$$\frac{\partial W}{\partial t_2} = \sum_{\alpha_1, \alpha_2=l}^3 \gamma_{\alpha_1\alpha_2} \frac{\partial}{\partial P_{\alpha_1}} \left( \beta \frac{P_{\alpha_2}}{m} + \frac{\partial}{\partial P_{\alpha_2}} \right) W, \quad (3.3.17)$$

where the tensorial friction coefficient  $\gamma_{\alpha_1\alpha_2}$  is given by

$$\gamma_{\alpha_1\alpha_2} \equiv \int_{-\infty}^0 d\sigma \left[ \left( f_{\alpha_1}(0) f_{\alpha_2}(\sigma) \right)^{Th} - f_{\alpha_1}^{Th} f_{\alpha_2}^{Th} \right]. \quad (3.3.18)$$

i.e.  $\gamma_{\alpha_1\alpha_2}$  is given by autocorrelation of the force on the nanoparticle. For anisotropic systems,

i.e. a virus migrating in a membranous cell interior,  $\gamma_{\alpha_1\alpha_2}$  is a second order tensor reflecting possible preferred migration directions.

Letting  $T = t_l$  we recompose the above results to obtain

$$\frac{\partial W}{\partial T} = -\frac{\bar{P}}{m} \cdot \frac{\partial W}{\partial \bar{R}} - \bar{f}^{Th} \cdot \frac{\partial W}{\partial \bar{P}} + \varepsilon^2 \bar{\gamma} \frac{\partial}{\partial \bar{P}} \left( \beta \frac{\bar{P}}{m} + \frac{\partial}{\partial \bar{P}} \right) W \quad (3.3.19)$$

in tensor notation.

Application of (3.3.19) to specific systems requires the use of Monte Carlo methods to estimate the friction coefficient and thermal average force. Note that the above results apply when the thermal average force evolves on the  $t_2$  scale, i.e. for an externally applied force that varies on a time scale much longer than that of atomic vibrations and collisions.

### 3.4 Collective Modes and Order Parameters for $N$ -atom Nanosystems

AMA of the Liouville equation starts with identification of a set of slow variables (i.e., order parameters). In hydrodynamics, conserved quantities are the traditional choice. For the present application, we suggest that they can be chosen according to the following criteria:

- A: they are expressible in terms of the  $6N$  atomic momenta and positions;
- B: they can be shown via Newton's equations to evolve on a timescale that is long compared to that of atomic vibrations or collisions;
- C: they capture the nanoscale phenomena of interest;
- D: the set of order parameters is complete, i.e. they are not strongly coupled to other order parameters not included in the model; and
- E: the energy of the system can be expressed in terms of these variables and residual rapidly-fluctuating atomistic variables.

As there are initial data or applied fields that can make any variable change rapidly, a corollary to criterion B is that the special conditions on problems of interest are met.

A nanoparticle in an aqueous medium has internal dynamics, e.g. the structural transitions 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 collective modes and order parameters that can be programmed to capture the salient

features of such systems and phenomena.

Principal component analysis (PCA) has been applied to reduce the dimensionality of MD trajectories for analyzing large-scale structural changes [115-117], it has been suggested to use PCA modes to characterize protein coherent motions and thus facilitate the simulation of long-time protein dynamics. Here we investigate its suitability for extracting our collective modes and order parameter as follows. PCA involves diagonalization of the positional covariance matrix for selected atoms (like  $C_\alpha$  in protein backbone), which generates an orthogonal set of eigenvectors (i.e., modes), denoted  $U \equiv \{\bar{u}_k, k = 1, \dots, 3N\}$ , each associated with an eigenvalue that indicates the amplitude of fluctuations along that mode. Eigenvalues divided by their sum represent the relative contributions of the corresponding modes to the total conformational change of the trajectory.

In order to use PCA modes obtained from short MD run to facilitate simulating long-time protein dynamics, it was hoped that the PCA modes with largest eigenvalues evolve slowly, characterizing coherent motions over long timescales, and those with small eigenvalues are high-frequency ones. Thus a few essential PCA modes can be chosen as a reduced representation for long-time protein dynamics. But its suitability has become controversial since the mid 1990s. It was noted in 1996 that due to the limited achievable sampling time of protein dynamics, PCA cannot extract properly the slowly-varying modes precluding it for dimensionality reduction [110]. In 2006, other authors arrived at the opposite conclusion [109] with one of the arguments being that it is more relevant to check the convergence of PCA subspaces (i.e., the summation of subspace modes weighted by their eigenvalues), rather than

that of individual modes. However, more studies showed that it is not reliable to use PCA modes for dimensionality reduction due to limited sampling of long-range correlations and forced orthogonalization of PCA modes [112-114].

To reexamine this issue, i.e., the suitability of using PCA for dimensionality reduction, we recently analyzed a 100ps ensemble-averaged MD trajectory of swollen CCMV capsid that was obtained by averaging 30 replica MD runs with random atomic velocity initializations at 298.15K [79]. By dividing the trajectory into 10 time windows, our calculations showed that PCA modes extracted from consecutive time windows display small similarities, which prove that they fail to capture system coherent motions. However, nanosystems are expected to undergo large-scale motions with atoms moving collectively and this is verified by the large correlations of the  $3N$ -dimensional atomic displacement vectors from consecutive time windows of the CCMV capsid trajectory.

In light of above findings, we construct collective modes to capture coherent motions of many-atom nanosystems by using atomic coordinates 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

$$\int_{-L/2}^{L/2} dx u_k(x) u_{k'}(x) = \delta_{kk'} \quad (3.4.1)$$

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

$$U_{\underline{k}}(\vec{s}) = u_{k_1}(x) u_{k_2}(y) u_{k_3}(z), \quad (3.4.2)$$

where  $\bar{s} = (x, y, z)$  for box centered at  $\bar{s} = \bar{0}$ . According to the space-warping method[118], a nanostructure embedded in a space ( $\bar{s}$  here) is considered to be a deformation of a reference space  $\bar{s}^0$ . The deformation of space is used to introduce order parameters via

$$\bar{s} = \sum_{\underline{k}} U_{\underline{k}}(\bar{s}^0) \bar{\Psi}_{\underline{k}}. \quad (3.4.3)$$

As the  $\bar{\Psi}_{\underline{k}}$  change,  $\bar{s}$ -space is deformed, and so is the nanosystem embedded in it. The  $\bar{\Psi}_{\underline{k}}$  constitute a set of vector order parameters that serve as the starting point of our AMA 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  $\bar{\Psi}_{\underline{k}}$ . However, given a finite truncation of the  $\underline{k}$  sum in (3.4.3), there will be some residual displacement of individual atoms that is not accounted for. Denoting this residual for atom  $i$  as  $\bar{\sigma}_i$ , we write

$$\bar{s}_i = \sum_{\underline{k}} U_{\underline{k}}(\bar{s}_i^0) \bar{\Psi}_{\underline{k}} + \bar{\sigma}_i. \quad (3.4.4)$$

This gives  $\bar{s}_i$  in the deformed state for the instantaneous values of the  $\bar{\Psi}_{\underline{k}}$  and  $\bar{\sigma}_i$ , and the reference configuration  $\bar{s}_i^0$ . If a sufficient number of terms are retained in the sum and the basis functions are chosen properly, then the  $\bar{\sigma}_i$  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  $\bar{\sigma}_i$  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  $\bar{s}_i^0 (i = 1, 2, \dots)$  could be regularly updated at a type of “piecewise continuation” in time by monitoring the  $\bar{\sigma}_i$  in the course of a simulation.

To evolve  $\bar{\Psi}_{\underline{k}}$  via Newtonian mechanics, and thereby start our multiscale analysis,  $\bar{\Psi}_{\underline{k}}$  must be expressed in terms of the  $\bar{s}_i$ . Assume  $m_i$  is the mass of particle  $i$  and  $m = \frac{1}{N} \sum_{i=1}^N m_i$ . We

determine  $\bar{\phi}_{\underline{k}}$  to minimize the mass-weighted root mean square residuals (RMSR), i.e.,

$\sqrt{\frac{1}{N} \sum_{i=1}^N \frac{m_i}{m} \bar{\sigma}_i^2}$ , which yields

$$\sum_{\underline{k}} B_{\underline{q}\underline{k}} \bar{\Psi}_{\underline{k}} = \frac{L^3}{N} \sum_{i=1}^N \frac{m_i}{m} U_{\underline{q}}(\bar{s}_i^0) \bar{s}_i, \quad B_{\underline{q}\underline{k}} = \frac{L^3}{N} \sum_{i=1}^N \frac{m_i}{m} U_{\underline{q}}(\bar{s}_i^0) U_{\underline{k}}(\bar{s}_i^0). \quad (3.4.5)$$

The  $\bar{\sigma}_i$  contribution is neglected in arriving at this definition of  $\bar{\Psi}_{\underline{k}}$  as  $\bar{\sigma}_i$  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, (3.4.5) is not an approximation; rather, the above discussion is a way to argue for the definition (3.4.5) of order parameters that captures coherent behaviors of a nanosystem. With this definition of  $\bar{\Psi}_{\underline{k}}$ , (3.4.5) is an exact relationship, since the  $\bar{\sigma}_i$  correct errors in the displaced atomic positions over-and-above the coherent contribution from the  $\bar{\Psi}_{\underline{k}}$  sum.

As we did not make any assumptions about the structure of the nanosystem for our derivations, there may be one nanoparticle in an aqueous host, multiple nanoparticles, viral capsids composed of protein capsomers, a virus and a cell membrane, or a liposome with its therapeutic payload. The ability to address composite nanosystems is a great benefit of the

present formulation. The orthogonality of the basis functions (3.4.1) implies that the matrix  $B$  is nearly diagonal and the order parameters 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  $B_{\underline{q}\underline{k}} \approx \delta_{\underline{q}\underline{k}}$  and (3.4.5) can be approximated as

$$\bar{\Psi}_{\underline{q}} \approx \frac{L^3}{N} \sum_{i=1}^N \frac{m_i}{m} U_{\underline{q}}(\bar{s}_i^0) \bar{s}_i. \quad (3.4.6)$$

In our implementation of the method, we provide both options to solve a linear system of equations in (3.4.5) for  $\bar{\Psi}_{\underline{k}}$  or compute them directly via (3.4.6). Inclusion of  $m_i$  in the above expressions gives the order parameters the character of generalized COM variables. Thus, if  $U_{\underline{k}}$  is constant then  $\bar{\Psi}_{\underline{k}}$  is proportional to the COM. Therefore, dimensionality reduction for nanostructures from the  $N$ -atom configuration to  $M$  order parameters is achieved through

$$\bar{s}_i, \{i = 1, \dots, N\} \xrightleftharpoons{\bar{u}_{\underline{k}}} \bar{\Psi}_{\underline{k}}, \{\underline{k} = \underline{k}_1, \dots, \underline{k}_M\}, \quad (3.4.7)$$

In the above formulation, atomic coordinates of a reference configuration are used to construct the nanostructure collective modes and their summation given by Eqn. (3.4.4) results in a new configuration that the system can evolve to. Instead atomic displacements calculated from two known configurations of the nanostructure can be used to construct its collective modes and their summations weighted by order parameters would describe system coherent motions directly [79]. Both of them can serve as reduced representation of the long-time dynamics of nanosystems. But for ease of implementation, the atomic positions of a reference



configuration are preferred, which avoids the need to find another system configuration to draw atomic displacements, even though it can be obtained through a short-time MD run or two states known to be available to the system (e.g., the native and swollen states for CCMV).

### 3.5 Free energy and friction coefficients

To implement AMA for viral structural transitions (STs), one needs to construct the free energy and friction coefficients for the range of slow variable values of interest. In this section, an approach for using constrained evolution to generate the nanocanonical ensemble for computing viral free energy and friction coefficients is presented and its computational feasibility of the approach is discussed.

Configurations and corresponding energies for fixed values of order parameters can be generated using the following non-inertial dynamics approach, i.e., constrained evolution for nanocanonical ensemble. The approach can also be generalized to generate the inertial dynamics needed to construct the friction coefficients. The ensemble of detailed atomic configurations for free energy computations are generated by solving the pseudo-dynamics equations

$$\frac{dr_{i\alpha}}{dt} = \sum_{i'\alpha'} B_{iai'\alpha'} \left[ -\frac{\partial V}{\partial r_{i'\alpha'}} + A_{i'\alpha'} \right], \quad (3.5.1)$$

where  $r_{i\alpha}$  is the  $\alpha$ -th position coordinate of atom  $i$ ,  $V$  is the  $N$ -atom potential and  $A_{i'\alpha'}$  is a random force. The  $B$  matrix is introduced to ensure that the  $\vec{r}_i$  only explore configurations consistent with given values of the order parameters. Consider a set of  $M$  order parameters

$\underline{\Phi} \equiv \{\Phi_k, k = 1, \dots, M\}$ . From each slow variable we construct the  $3N$  length column vectors  $\partial\Phi_k/\partial r_{i\alpha}, i = 1, \dots, N; \alpha = 1, 2, 3$ . From these we form a set of  $M$  orthonormalized row  $\langle k|$  and column  $|k\rangle$  vectors and thereby the projection matrix  $|k\rangle\langle k|$ . Then, in matrix notation,  $\underline{B} = \underline{I} - |1\rangle\langle 1| - \dots - |M\rangle\langle M|$ . With this and (3.5.1), one finds that  $d\Phi_k/dt = 0, k = 1, \dots, M$ . The ensemble generation can be stabilized by replacing  $V$  with  $U(\underline{\Phi}; \bar{r}_1, \dots, \bar{r}_N)$  defined in Sect. 3.1. Thus (3.5.1) drives the system to a state with  $\underline{\Phi}$ -dependent minimum and there is no gradual drift of the  $\underline{\Phi}$  due to numerical round-off. The ensembles so generated can be enriched by using multiple initial datasets for the  $r_{i\alpha}$  to avoid being trapped in a local minimum of  $U$ . Space-warping can be used to generate these configurations as it facilitates transitions between configurations with significantly different properties. The utility of this approach in minimization has already been demonstrated[118]. The ensemble for each choice of the order parameters can be used as a basis of a Monte Carlo integration algorithm to compute the thermal average forces needed for the FP equation.

As reviewed in Sect. 2.1, an all-atom molecular dynamics simulation on a complete satellite tobacco mosaic virus has been achieved on a SGI 1,024-processor Altix system by using 256 processors and 128 GB memory with a performance of 1.1 ns/day ( $7.85 \times 10^{-2}$  sec/time step)[52]. We believe that it should take about 10000 such cycles with intermittent major moves using our space-warping technique to obtain a free energy variation with 20 wisely-chosen slow variable values. This would take 4.36 hours on a computer platform with similar performance to the Altix system noted above. A more complete temporal dynamics will be used in a similar way to construct thermal average forces and friction coefficients. These computations will benefit from

published methods [100, 119-122].

### 3.6 Generalizations

The AMA methodology can be generalized to complex systems involving nanoparticles with complex internal structure or the interaction of multiple nanoparticles. For example, due to its size, a nanoparticle may experience phase transition-like behaviors. In this case the associated order parameters, COM position and orientation constitute a set of order parameters denoted  $\underline{R}$ ; with the latter are conjugate momenta  $\underline{P}$ . This theme can be extended to many nanoparticle systems and complex host media like a cell's interior. As for the COM variables of Sects. 3.2 and 3.3, one can introduce a nanocanonical ensemble such that

$$\hat{\rho}_{\underline{M}, \underline{K}} = \frac{\exp(-\beta H_0 + \underline{M} \cdot \underline{P} + \underline{K} \cdot \underline{R})}{\Xi(\beta, \underline{M}, \underline{K})} \quad (3.6.1)$$

where

$$\Xi(\beta, \underline{M}, \underline{K}) = \int d^{6N} \Gamma e^{-\beta H_0 + \underline{M} \cdot \underline{P} + \underline{K} \cdot \underline{R}}. \quad (3.6.2)$$

The transformation  $\Psi/\Xi \rightarrow W/Q$  follows directly as in Sect. 3.2, as does the removal of secular behavior without atomic momentum-position integration as in Sect. 3.3. The result of the multiscale computations is a generalized Fokker-Planck equation in the inertial limit. While computations without the present approach would require applying  $\int d^{6N} \Gamma$  but with integrations involving complex restrictions (stemming from the many  $\underline{R}, \underline{P}$  variables that must be respected in order to conserve the number of degrees of freedom).

As general collective modes and order parameters are systematically constructed in Sect. 3.4

to capture the coherent motions of nanosystems, they can be taken as input for AMA of the  $N$ -atom Liouville equation. This may need the analysis of the scaled Liouville equation to orders higher than  $\varepsilon^4$  and lead to stochastic equations for the dynamics of the system collective modes and order parameters, which provide more complete, general multiscale models for describing the structural transitions of viruses and other nanosystems. Note that our above AMA formulation was derived in the inertial limit for order parameters, for which a FP equation was obtained. In the friction-dominant case, AMA of the Liouville equation could result in a Smoluchowski equation for the time evolution of order parameters [102, 123, 124].

## Chapter 4 Transition State Theory for Viral Dilatational Structural Transitions

### 4.1 Multiscale Viral Dilational Structural Transition Model

Following from the AMA formulation presented in Chapter 3, order parameters capturing viral dilatation are constructed below and a Fokker-Planck (FP) equation for their long timescale dynamics is derived. First denote the COM position of the virus  $\bar{R}^*$ :

$$\bar{R}^* = \sum_{i=1}^N \frac{m_i \bar{r}_i}{m^*} \Theta_i, \quad m^* = \sum_{i=1}^N m_i \Theta_i, \quad (4.1.1)$$

where  $\Theta_i = 1$  if atom  $i$  is in the virus and 0 otherwise, and  $m^*$  is the total viral mass.

Newton's equations imply that  $d\bar{R}^*/dt = -\mathcal{L}\bar{R}^*$ , in which  $\mathcal{L}$  is the Liouville operator

$$\mathcal{L} = -\sum_{i=1}^N \left[ \frac{\bar{p}_i}{m_i} \cdot \frac{\partial}{\partial \bar{r}_i} + \bar{F}_i \cdot \frac{\partial}{\partial \bar{p}_i} \right], \quad (4.1.2)$$

where  $\bar{F}_i = -\left( \frac{\partial V}{\partial \bar{r}_i} \right)_{\bar{r}_{j \neq i}}$  is the force on atom  $i$  and  $V(\bar{r}_1, \dots, \bar{r}_N)$  is the  $N$ -atom potential. With

this and introducing the total viral momentum  $\bar{P}^*$  we obtain  $d\bar{R}^*/dt = \bar{P}^*/m^*$  for

$$\bar{P}^* = \sum_{i=1}^N \bar{p}_i \Theta_i. \text{ Similarly, } d\bar{P}^*/dt = -\mathcal{L}\bar{P}^* = \sum_{i=1}^N \bar{F}_i \Theta_i \text{ (the net force on the virus).}$$

The virus contains many atoms and hence has large size relative to that of an atom; thus we take it to have diameter of  $O(\varepsilon^{-1})$  for small parameter  $\varepsilon$ . Since we are interested in significant migration distances (i.e. greater than or equal to the viral diameter), we scale  $\bar{R}^*$  to be  $O(\varepsilon^{-1})$ . The scaling  $\varepsilon^2 = m/m^*$  for typical atomic mass  $m$  is adopted under the

assumption that the virus is empty—i.e. we develop a theory for the dynamics of a viral capsid. Other cases may easily be considered as well (e.g.  $m^*$  is  $O(\varepsilon^{-3})$  for a complete virus with genome in the capsid's cavity). Under the assumption that the system is near equilibrium, a typical viral COM kinetic energy  $P^{*2}/2m^*$  is  $O(k_B T)$ . Thus  $\bar{P}^*$  scales as  $\sqrt{m^*}$ , i.e. is  $O(\varepsilon^{-1})$ . With this we adopt the scaled variables  $\bar{R}$  and  $\bar{P}$  such that

$$\bar{R}^* = \varepsilon^{-1} \bar{R}, \quad \bar{P}^* = \varepsilon^{-1} \bar{P}. \quad (4.1.3)$$

Let  $\bar{f} = \varepsilon^{-1} \sum_{i=1}^N \bar{F}_i \Theta_i$  be the scaled net force on the virus; this scaling emerges from the near-equilibrium assumption and that the viral diameter is  $O(\varepsilon^{-1})$  and hence its surface area is  $O(\varepsilon^{-2})$ ; for short range interactions with the host medium the number of contributions to the net force on the virus is proportional to its surface area; however near equilibrium there is much cancellation of these forces so that the residual force over-and-above its zero average is small and hence scaled to be on the order of one over the square root of the number of individual atomic contributions, i.e.  $O(\varepsilon)$ . This is one example of the corollary to criterion B, i.e. the net force is sufficiently small that the momentum evolves slowly. In contrast, if the host medium is experiencing a shock wave, the number of host atoms on one side of the virus would be much larger than on the other and hence  $\bar{P}$  would change rapidly as the shock is passing. In summary, under the above assumptions Newton's equations imply

$$\frac{d\bar{R}}{dt} = \varepsilon^2 \frac{\bar{P}}{m}, \quad \frac{d\bar{P}}{dt} = \varepsilon^2 \bar{f}, \quad (4.1.4)$$

proving that  $\bar{R}$  and  $\bar{P}$  are slowly varying.

To characterize intraviral structural dynamics, we introduce relative coordinates  $\bar{s}_i$  for viral atoms such that

$$\bar{r}_i = \bar{s}_i + \Theta_i \bar{R}^*. \quad (4.1.5)$$

Next an order parameter  $\Phi$  is introduced to describe viral dilatation. In particular we take  $\Phi$  to be a measure of the viral size relative to that of a reference structure, e.g. an X-ray crystal structure. To this end consider the definition

$$\Phi = \frac{\sum_{i=1}^N m_i \bar{s}_i \cdot \bar{\bar{X}} \bar{s}_i^0 \Theta_i}{\sum_{i=1}^N m_i s_i^{02} \Theta_i}, \quad (4.1.6)$$

where  $\bar{\bar{X}}$  is a length-preserving rotation matrix (see below) and  $s_i^0 = |\bar{s}_i^0|$ . We now demonstrate that this order parameter and the associated momentum  $\Pi$  are order parameters. This  $\Phi, \Pi$  order parameter description will serve as the basis for a simple viral dilatation phase transition theory.

The equation of motion of  $\Phi$  is

$$\frac{d\Phi}{dt} = -\mathcal{L}\Phi = \frac{\varepsilon^4}{m'} \sum_{i=1}^N \left( \frac{\bar{p}_i}{m_i} - \frac{\bar{P}^*}{m^*} \right) \cdot \bar{\bar{X}} \bar{s}_i^0 m_i \Theta_i, \quad m' \equiv \varepsilon^4 \sum_{i=1}^N m_i s_i^{02} \Theta_i. \quad (4.1.7)$$

The scaling of  $m'$  is consistent with our earlier assumption that the number of atoms in the virus

is  $O(\varepsilon^{-2})$ , and its diameter is  $O(\varepsilon^{-1})$ . Introducing relative velocities  $\frac{\bar{\pi}_i}{m_i} = \frac{\bar{p}_i}{m_i} - \frac{\bar{P}^*}{m^*}$  for viral

atoms yields

$$\frac{d\Phi}{dt} = -\mathcal{L}\Phi = \varepsilon^2 \frac{\Pi}{m'}, \quad \Pi = \varepsilon^2 \Pi^*, \quad \Pi^* \equiv \sum_{i=1}^N \vec{\pi}_i \cdot \vec{\bar{X}} \vec{s}_i^0 \Theta_i. \quad (4.1.8)$$

The scaling of  $\Pi$  is based on the assumption that while there are  $O(\varepsilon^{-2})$  atoms in the virus, the contributions to  $\Pi^*$  are of fluctuating sign, but each term has an  $\vec{s}_i^0$  factor which, like the viral diameter, is  $O(\varepsilon^{-1})$ ; thus  $\Pi^*$  is  $O(\varepsilon^{-2})$ . With this

$$\frac{d\Pi}{dt} = -\mathcal{L}\Pi = \varepsilon^2 \sum_{i=1}^N \vec{F}_i \cdot \left( \vec{\bar{X}} \vec{s}_i^0 - \sum_{j=1}^N \frac{m_j \vec{\bar{X}} \vec{s}_j^0}{m^*} \Theta_j \right) \Theta_i. \quad (4.1.9)$$

The  $j$ -sum in (4.1.9) is over a large number of vector contributions which tend to cancel; as the  $m^*$  factor in this term is proportional to the number of atoms in the virus, the  $j$ -sum is small relative to  $\vec{\bar{X}} \vec{s}_i^0$ . Thus to good approximation we rewrite (4.1.9) as

$$\frac{d\Pi}{dt} = \varepsilon^2 g, \quad g = \sum_{i=1}^N \vec{F}_i \cdot \vec{\bar{X}} \vec{s}_i^0 \Theta_i, \quad (4.1.10)$$

for “dilatation force”  $g$  that is the analogue of the COM force  $\vec{f}$ . The scaling of  $g$  as implied in (4.1.10) differs from that of  $\vec{f}$  as each term in  $g$  has an additional  $\vec{s}_i^0$  factor (which is  $O(\varepsilon^{-1})$ ).

With this we conclude that  $\Phi$  and  $\Pi$  are also order parameters.

Viral rotation can also be shown to be slow due to the large moment of inertia. First relate the rotation matrix  $\vec{\bar{X}}$  to the relative atomic configuration  $\{\vec{s}_i, \Theta_i = I\}$ . If the intraviral state only changes due to overall rotation and isotropic dilatation, then the  $\alpha_2$  component of  $\vec{s}_i$  is given by

$$s_{i\alpha_2} = \Phi \sum_{\alpha_1=1}^3 X_{\alpha_2\alpha_1} s_{i\alpha_1}^0. \quad (4.1.11)$$



Multiplying by  $s_{i\alpha_3}^0 \Theta_i$  and summing over  $i$  yields

$$Y_{\alpha_3\alpha_2} = \Phi \sum_{\alpha_1=1}^3 X_{\alpha_2\alpha_1} Y_{\alpha_3\alpha_1}^0, \quad (4.1.12)$$

$$Y_{\alpha_3\alpha_2} = \sum_{i=1}^N s_{i\alpha_3}^0 s_{i\alpha_2} \Theta_i, \quad Y_{\alpha_3\alpha_1}^0 = \sum_{i=1}^N s_{i\alpha_3}^0 s_{i\alpha_1}^0 \Theta_i. \quad (4.1.13)$$

Taking  $d/dt$  of both sides of (4.1.12) and recalling that  $\bar{s}_i^0$  is a reference configuration (and not a dynamical variable) yields

$$\sum_{i=1}^N \frac{s_{i\alpha_3}^0 \pi_{i\alpha_2} \Theta_i}{m_i} = \Phi \sum_{\alpha_1=1}^3 \frac{dX_{\alpha_2\alpha_1}}{dt} Y_{\alpha_3\alpha_1}^0 + \sum_{\alpha_1=1}^3 X_{\alpha_2\alpha_1} Y_{\alpha_3\alpha_1}^0 \frac{d\Phi}{dt}. \quad (4.1.14)$$

Since the virus diameter is  $O(\varepsilon^{-1})$ , then so are  $\bar{s}_i$  and  $\bar{s}_i^0$  for viral atoms. This implies that  $Y_{\alpha_3\alpha_1}^0$  is  $O(\varepsilon^{-4})$  as it is a sum of  $O(\varepsilon^{-2})$  positive terms (at least for the diagonal elements). Consider the trace of the LHS of (4.1.14), i.e. the sum of  $\bar{s}_i^0 \cdot \bar{\pi}_i / m_i$  terms. Using an argument as for  $\Pi^*$ , it is expressed to be  $O(\varepsilon^{-2})$ . With this, and that  $d\Phi/dt$  is  $O(\varepsilon^2)$ ,  $d\bar{\bar{X}}/dt$  is seen to be  $O(\varepsilon^2)$  so that rotation is slow.

While more order parameters characterizing the viral internal structure need to be explored apart from their COM position, orientation, dilatational order parameter and their associated momenta as constructed above, the AMA formulation of the  $N$ -atom Liouville equation will be demonstrated with above order parameters ( $\bar{P}$ ,  $\bar{R}$ ,  $\Pi$ ,  $\Phi$ ,  $\bar{\bar{X}}$  and  $\bar{\bar{Q}}$ ) to investigate STs of icosahedral viruses. Since icosahedral viruses are nearly spherical, this symmetry and the conjecture that centrifugal forces do not have significant effect on STs, we expect that effects of overall rotation (i.e.  $\bar{\bar{X}}$  and  $\bar{\bar{Q}}$ ) should decouple from those of STs. Hence the reduced

distribution for the order parameters  $W(\bar{P}, \bar{R}, \Pi, \Phi, \underline{t})$  can be factorized into  $\bar{\bar{X}}$ ,  $\bar{\bar{\Omega}}$ , and  $\bar{P}$ ,  $\bar{R}$ ,  $\Pi$ ,  $\Phi$  dependent parts. With this, the  $N$ -atom probability density  $\rho$  is taken to have the dependence  $\rho(\Gamma; \bar{P}, \bar{R}, \Pi, \Phi, t_0, \underline{t})$  with  $\Gamma \equiv \{\bar{p}_1, \bar{r}_1, \dots, \bar{p}_N, \bar{r}_N\}$  while  $\bar{\bar{X}}$  and  $\bar{\bar{\Omega}}$  are ignored henceforth.

## 4.2 Order Parameter Dependences of System Energy and Lowest Order Probability Density

To complete the multiscale analysis, the  $N$ -atom potential  $V$  and kinetic energy  $K$  must be expressed in terms of slow coordinates ( $\bar{R}$ ,  $\Phi$ ,  $\bar{\bar{X}}$  for viral dilatational STs), the residual dependence on  $\bar{r}_1, \dots, \bar{r}_N$ , and the associated momenta. To accomplish this for  $V$  we write

$$\bar{r}_i = \bar{\sigma}_i + \Theta_i \left( \varepsilon^{-1} \bar{R} + \Phi \bar{\bar{X}}_i^0 \right). \quad (4.2.1)$$

The  $\bar{R}^*$  term accounts for migration of the viral COM, the  $\Phi \bar{\bar{X}}$  contribution generates an atomic displacement due to rotation and dilatation, and  $\bar{\sigma}_i$  is the residual (incoherent) displacement over-and-above the coherent effects of  $\bar{R}$ ,  $\Phi$ , and  $\bar{\bar{X}}$ . With this we introduce the explicit potential function  $U$  via

$$U(\bar{R}, \Phi, \bar{\bar{X}}; \bar{r}_1, \dots, \bar{r}_N) = V \left( \left\{ \bar{\sigma}_i + \Theta_i \left( \varepsilon^{-1} \bar{R} + \Phi \bar{\bar{X}}_i^0 \right); i = 1, \dots, N \right\} \right). \quad (4.2.2)$$

The gradient of  $V$  with respect to  $\Phi$  is given by the derivative of  $U$  with respect to  $\Phi$  keeping  $\bar{R}$ ,  $\bar{\bar{X}}$ , and the residual  $\bar{r}_i$ -dependence fixed; this shows the utility in introducing  $U$ , i.e. the residual dependence on the  $\bar{r}_i$  from the  $\bar{\sigma}_i$  is explicit. With this the scaled net force on the virus  $\bar{f}$  and dilatation force  $g$  are given by

$$\vec{f} = \varepsilon^{-1} \sum_{i=1}^N \vec{F}_i \Theta_i = - \left( \frac{\partial U}{\partial \vec{R}} \right)_{\Phi, \vec{X}, \vec{r}_1, \dots, \vec{r}_N}, \quad (4.2.3)$$

$$g = - \left( \frac{\partial U}{\partial \Phi} \right)_{\vec{R}, \vec{X}, \vec{r}_1, \dots, \vec{r}_N} = \sum_{i=1}^N \vec{F}_i \cdot \vec{\bar{X}} \vec{s}_i^0 \Theta_i. \quad (4.2.4)$$

The kinetic energy  $K$  of the virus is given by

$$K = \sum_{i=1}^N \frac{p_i^2}{2m_i} \Theta_i. \quad (4.2.5)$$

Letting  $\vec{p}_i^{inc}/m_i = -\mathcal{L}\vec{\sigma}_i$  and  $\vec{p}_i^{coh}/m_i = -\mathcal{L}(\vec{r}_i - \vec{\sigma}_i)$ , we obtain

$$K = \sum_{i=1}^N \frac{1}{2m_i} \left[ \left( p_i^{inc} \right)^2 + 2 \vec{p}_i^{inc} \cdot \vec{p}_i^{coh} + \left( p_i^{coh} \right)^2 \right] \Theta_i. \quad (4.2.6)$$

Consider the coh/inc cross-term; for a small cluster (say 100) of the million atoms in the virus.

The  $\vec{p}_i^{coh}$  in this cluster are in the same direction but the  $\vec{p}_i^{inc}$  will have varying directions. Even

if there is a small residual contribution from the cross-term of one cluster, it is likely that there is

another to cancel it. With this we take the cross-term to be negligible.

Letting  $\varepsilon^2 \vec{\bar{\Omega}} = -\mathcal{L}\vec{\bar{X}}$ , we find that

$$\vec{p}_i^{coh} = m_i \Theta_i \left[ \varepsilon \frac{\vec{P}}{m} + \varepsilon^2 \frac{\Pi}{m'} \vec{\bar{X}} \vec{s}_i^0 + \varepsilon^2 \Phi \vec{\bar{\Omega}} \vec{s}_i^0 \right]. \quad (4.2.7)$$

Note that the  $\Pi$  contribution corresponds to motion along the  $\vec{\bar{X}} \vec{s}_i^0$  direction while the  $\vec{\bar{\Omega}}$

contribution is a rotation about the COM; thus the  $\Pi$  and  $\vec{\bar{\Omega}}$  contributions are orthogonal.

With this, and recalling that  $\left| \vec{\bar{X}} \vec{s}_i^0 \right| = s_i^0$ , we obtain

$$K^{coh} = \sum_{i=1}^N \frac{(p_i^{coh})^2}{2m_i} \Theta_i = \frac{1}{2} \sum_{i=1}^N m_i \Theta_i \left[ \varepsilon^2 \frac{P^2}{m^2} + \varepsilon^4 \frac{\Pi^2}{m'^2} s_i'^{02} + \varepsilon^4 \Phi^2 \left| \bar{\bar{\Omega}} \bar{s}_i'^0 \right|^2 \right]. \quad (4.2.8)$$

Cross-terms of  $\bar{P}$  with  $\Pi$  or  $\bar{\bar{\Omega}}$  are dropped as the latter contributions are in many directions so that the  $i$ -sum makes them negligible. Note that  $K^{coh}$  is then independent of the residual variations in  $\Gamma$  (i.e. those over-and-above that in  $\bar{P}$ ,  $\Pi$  and  $\bar{\bar{\Omega}}$ ) so that the partition function  $Q$  (see below) can be written as a configuration part (i.e. one dependent on  $\bar{R}$ ,  $\Phi$  and  $\bar{\bar{X}}$ ) times  $\exp(-\beta K^{coh})$ . The incoherent part of  $K$  can be written as

$$K^{inc} = \sum_{i=1}^N \frac{(p_i^{inc})^2}{2m_i} \Theta_i = \sum_{i=1}^N \frac{\Theta_i}{2m_i} \left| \bar{p}_i - \bar{p}_i^{coh} \right|^2. \quad (4.2.9)$$

In the lowest order solution to the scaled Liouville equation as analyzed in Sect. 4.1, we have the following factor

$$\hat{\rho} = \frac{e^{-\beta(K^{inc} + V)}}{Q^{inc}(\beta, \bar{R}, \Phi, \bar{\bar{X}})}, \quad (4.2.10)$$

$$Q^{inc} = \int d^{6N} \Gamma \Delta \exp \left[ -\beta(K^{inc} + V) \right], \quad (4.2.11)$$

where  $\Delta$  is a product of delta functions of  $\bar{R}$ ,  $\Phi$  and  $\bar{\bar{X}}$  centered around their  $\Gamma$ -dependent values. As  $\bar{p}_i^{inc} = \bar{p}_i - \bar{p}_i^{coh}$  and  $\bar{p}_i^{coh}$  is  $O(\varepsilon)$ ,  $\partial \hat{\rho} / \partial \bar{P}$ ,  $\partial \hat{\rho} / \partial \Pi$  and  $\partial \hat{\rho} / \partial \bar{\bar{\Omega}}$  are zero.

### 4.3 Derivation of the Fokker-Planck equation of Order Parameters

With the  $N$ -atom probability density  $\rho$  taken to have the dependence  $\rho(\Gamma; \bar{P}, \bar{R}, \Pi, \Phi, t_0, \underline{t})$  with  $\Gamma \equiv \{\bar{p}_1, \bar{r}_1, \dots, \bar{p}_N, \bar{r}_N\}$ , we introduce a set of scaled times  $t_n = \varepsilon^{2n} t, n = 0, 1, \dots$ . With this and the chain rule, the Liouville equation takes the form

$$\sum_{n=0}^{\infty} \varepsilon^{2n} \frac{\partial \rho}{\partial t_n} = (\mathcal{L}_0 + \varepsilon^2 \mathcal{L}_1) \rho \quad (4.3.1)$$

$$\mathcal{L}_0 = - \sum_{i=1}^N \left[ \frac{\bar{p}_i}{m_i} \cdot \frac{\partial}{\partial \bar{r}_i} + \bar{F}_i \cdot \frac{\partial}{\partial \bar{p}_i} \right] \quad (4.3.2)$$

$$\mathcal{L}_1 = - \frac{\bar{P}}{m} \cdot \frac{\partial}{\partial \bar{R}} - \bar{f} \cdot \frac{\partial}{\partial \bar{P}} - \frac{\Pi}{m'} \frac{\partial}{\partial \Phi} - g \frac{\partial}{\partial \Pi} . \quad (4.3.3)$$

The operators  $\mathcal{L}_0$  and  $\mathcal{L}_1$  arise from the chain rule in the course of accounting for the direct and indirect dependence of  $\rho$  on  $\Gamma$ ; therefore these operators only have meaning when acting on  $\rho$  in its present form as a function of both  $\Gamma$  and the order parameters.

At first sight one might suggest choosing the relative variables  $\bar{\pi}_i, \bar{s}_i$  of Sect. II for the viral atoms to express  $\mathcal{L}_0$  and  $\mathcal{L}_1$  and develop the theory, rather than the  $\bar{p}_i, \bar{r}_i$  defined in the laboratory frame as used here. However,  $\sum_{i=1}^N \bar{\pi}_i \Theta_i = \bar{0}$  and  $\sum_{i=1}^N m_i \bar{s}_i \Theta_i = \bar{0}$ . This implies that the  $\{\bar{\pi}_i, \bar{s}_i; \Theta_i = I\}$  do not constitute a set of independent variables. In contrast, the  $\bar{p}_i, \bar{r}_i$  are independent. While the order parameters depend on  $\Gamma$ , the partial derivatives of  $\rho$  with respect to the  $\bar{p}_i, \bar{r}_i$  at constant order parameters in  $\mathcal{L}_0$ , and conversely for derivatives in  $\mathcal{L}_1$ , only reflect the use of the chain rule to account for the direct and indirect dependencies of  $\rho$  on  $\Gamma$ . It does not constitute a validation of the number of degrees of freedom. In contrast, ignoring the fact that the  $\bar{\pi}_i$  and  $\bar{s}_i$  are not independent variables would introduce a violation of the number of degrees of freedom.

Development of the theory proceeds by constructing a perturbative solution for  $\rho$ , i.e.

$$\rho = \sum_{n=0}^{\infty} \rho_n \varepsilon^{2n}. \quad (4.3.4)$$

To lowest order, with the assumption that  $\rho_0$  is independent of the microscopic time  $t_0$ , we arrive at  $\mathcal{L}_0 \rho_0 = 0$ ; this implies that the biological phenomena of interest have quasi-equilibrium character (e.g. we are not interested in processes on the  $10^{-12}$  sec timescale). The nanocanonical quasi-equilibrium solution of this equation is found to be <sup>35</sup>

$$\rho_0 = \frac{e^{-\beta H_0}}{Q(\beta, \bar{R}, \Phi)} W(\bar{P}, \bar{R}, \Pi, \Phi, \underline{t}) \equiv \hat{\rho} W. \quad (4.3.5)$$

This illustrate the multiple dependencies on  $\Gamma$ , i.e. direct and indirect through  $H_0$ , and indirectly through the order parameters appearing in  $Q$  and  $W$ . The reduced distribution  $W$  can be shown to be the average of the microscopic density  $\delta(\bar{p} - \bar{P})\delta(\bar{r} - \bar{R})\delta(\pi - \Pi)\delta(\phi - \Phi)$  for the lowest order  $N$ -atom distribution  $\rho_0$ . The variables  $\bar{p}, \bar{r}, \pi, \phi$  are particular values of the dynamical variables  $\bar{P}, \bar{R}, \Pi, \Phi$  but are not dynamical variables (i.e.  $\Gamma$ -dependent quantities themselves). In contrast,

$$\bar{P} \equiv \varepsilon \sum_{i=1}^N \bar{p}_i \Theta_i, \bar{R} \equiv \varepsilon \sum_{i=1}^N \frac{m_i \bar{r}_i}{m^*} \Theta_i, \Pi \equiv \varepsilon^2 \sum_{i=1}^N \bar{\pi}_i \cdot \bar{\bar{X}} \bar{s}_i^0 \Theta_i, \Phi \equiv \frac{\sum_{i=1}^N m_i \bar{s}_i \cdot \bar{\bar{X}} \bar{s}_i^0 \Theta_i}{\sum_{i=1}^N m_i s_i^{02} \Theta_i}. \quad (4.3.6)$$

The  $\bar{r}, \phi$ -dependent partition function  $Q$  is given by

$$Q(\beta, \bar{r}, \phi) = \int d^{6N} \Gamma \delta(\bar{r} - \bar{R}) \delta(\phi - \Phi) e^{-\beta H_0}, \quad (4.3.7)$$

$$H_0 = \sum_{i=1}^N \frac{(p_i^{inc})^2}{2m_i} \Theta_i + U(\bar{R}, \Phi; \bar{r}_1, \dots, \bar{r}_N), \quad (4.3.8)$$

where  $\bar{p}_i^{inc}$  is defined in terms of the  $\bar{p}_i$  and the order parameters of (4.3.6) in (A.9).

To  $O(\varepsilon^2)$  the Liouville equation implies

$$\left(\frac{\partial}{\partial t} - \mathcal{L}_0\right)\rho_1 = -\frac{\partial \rho_0}{\partial t_1} + \mathcal{L}_1 \rho_0, \quad (4.3.9)$$

where we drop the 0 on for simplicity henceforth. Assuming that initially  $\rho$  is near-equilibrium,  $\rho_1$  can be taken to be zero at  $t=0$ . With this, the  $O(\varepsilon^2)$  equation admits the solution

$$\rho_1 = \int_0^t dt' e^{\mathcal{L}_0(t-t')} \left[ -\tilde{\rho} \frac{\partial W}{\partial t_1} + \mathcal{L}_1 \rho W \right]. \quad (4.3.10)$$

Applying  $\mathcal{L}_1$  and noting that  $\mathcal{L}_0 \hat{\rho} = 0$  yields

$$\rho_1 = -t \tilde{\rho} \frac{\partial W}{\partial t_1} - \int_0^t dt' e^{\mathcal{L}_0(t-t')} \left[ \frac{\bar{P}}{m} \cdot \frac{\partial}{\partial \bar{R}} + \bar{f} \cdot \frac{\partial}{\partial \bar{P}} + \frac{\Pi}{m'} \frac{\partial}{\partial \Phi} + g \frac{\partial}{\partial \Pi} \right] \rho W. \quad (4.3.11)$$

The statistical mechanical postulate “the longtime and ensemble averages for equilibrium systems are equal” implies

$$\lim_{t \rightarrow \infty} \frac{1}{t} \int_{-t}^0 dt'' e^{-\mathcal{L}_0 t''} A = \int d^{6N} \Gamma A \hat{\rho} \equiv A^{th} \quad (4.3.12)$$

for any dynamical variable  $A(\Gamma)$ . Changing variables via  $t'' = t' - t$  in (4.3.11), it is found that removal of the secular behavior in  $\rho_1$  at large  $t$  implies

$$\frac{\partial W}{\partial t_1} = - \left[ \frac{\bar{P}}{m} \cdot \frac{\partial}{\partial \bar{R}} + \bar{f}^{th} \cdot \frac{\partial}{\partial \bar{P}} + \frac{\Pi}{m'} \frac{\partial}{\partial \Phi} + g^{th} \frac{\partial}{\partial \Pi} \right] W, \quad (4.3.13)$$

where  $\bar{f}^{th}$  and  $g^{th}$  are the  $\hat{\rho}$ -weighted thermal average quantities and the equivalence of long

time and thermal averages have been assumed. With this

$$\rho_1 = -\hat{\rho} \int_0^t dt' e^{\mathcal{L}_0(t-t')} \left\{ (\bar{f} - \bar{f}^{th}) \cdot \left( \beta \frac{\bar{P}}{m} + \frac{\partial}{\partial \bar{P}} \right) + (g - g^{th}) \left( \beta \frac{\Pi}{m'} + \frac{\partial}{\partial \Pi} \right) \right\} W. \quad (4.3.14)$$

Note that the use of the long time/ensemble average equivalence avoids the traditional use of the integration of the Liouville equation over  $\Gamma$  which both erases ambiguities (i.e. do not include the  $\Gamma$ -dependence in the order parameters) and insures that all secular behavior in  $\rho$  is removed, not just that in a reduced distribution. This completes the  $O(\varepsilon^2)$  analysis.

To  $O(\varepsilon^4)$  one finds

$$\left( \frac{\partial}{\partial t} - \mathcal{L}_0 \right) \rho_2 = -\frac{\partial \rho_0}{\partial t_2} - \frac{\partial \rho_1}{\partial t_1} + \mathcal{L}_1 \rho_1. \quad (4.3.15)$$

Again with  $\rho_2 = 0$  at  $t = 0$ , we find

$$\rho_2 = \int_0^t dt' e^{\mathcal{L}_0(t-t')} \left( -\hat{\rho} \frac{\partial W}{\partial t_2} - \frac{\partial \rho_1}{\partial t_1} + \mathcal{L}_1 \rho_1 \right). \quad (4.3.16)$$

Removal of the secular behavior in  $\rho_2$  at large  $t$  implies

$$\begin{aligned} \frac{\partial W}{\partial t_2} = & \left[ \bar{\gamma}_{ff} \frac{\partial}{\partial \bar{P}} \left( \beta \frac{\bar{P}}{m} + \frac{\partial}{\partial \bar{P}} \right) + \bar{\gamma}_{fg} \cdot \frac{\partial}{\partial \bar{P}} \left( \beta \frac{\Pi}{m'} + \frac{\partial}{\partial \Pi} \right) \right. \\ & \left. + \bar{\gamma}_{gf} \cdot \frac{\partial}{\partial \Pi} \left( \beta \frac{\bar{P}}{m} + \frac{\partial}{\partial \bar{P}} \right) + \gamma_{gg} \frac{\partial}{\partial \Pi} \left( \beta \frac{\Pi}{m'} + \frac{\partial}{\partial \Pi} \right) \right] W. \end{aligned} \quad (4.3.17)$$

The friction tensors are given by



$$\begin{aligned}
\gamma_{ff,kl} &\equiv \int_{-\infty}^0 dt \left[ \left( f_k(0) f_l(t) \right)^{th} - f_k^{th} f_l^{th} \right], \\
\gamma_{fg,k} &\equiv \int_{-\infty}^0 dt \left[ \left( f_k(0) g(t) \right)^{th} - f_k^{th} g^{th} \right], \\
\gamma_{gf,k} &\equiv \int_{-\infty}^0 dt \left[ \left( g(0) f_k(t) \right)^{th} - g^{th} f_k^{th} \right], \\
\gamma_{gg} &\equiv \int_{-\infty}^0 dt \left[ \left( g(0) g(t) \right)^{th} - g^{th} g^{th} \right], \\
k, l &= 1, 2, 3.
\end{aligned} \tag{4.3.18}$$

Thus the friction coefficients are related to force correlation functions as expected.

Letting  $\tau = t_l$  we recompose the above results<sup>43</sup> to obtain a FP equation for the reduced distribution of viral order parameters  $W(\bar{P}, \bar{R}, \Pi, \Phi, \underline{t})$  as:

$$\begin{aligned}
\frac{\partial W}{\partial \tau} = & - \left[ \frac{\bar{P}}{m} \frac{\partial}{\partial \bar{R}} + \bar{f}^{th} \frac{\partial}{\partial \bar{P}} + \frac{\Pi}{m'} \frac{\partial}{\partial \Phi} + g^{th} \frac{\partial}{\partial \Pi} \right] W + \\
& \varepsilon^2 \left[ \bar{\gamma}_{ff} \frac{\partial}{\partial \bar{P}} \left( \beta \frac{\bar{P}}{m} + \frac{\partial}{\partial \bar{P}} \right) + \bar{\gamma}_{fg} \cdot \frac{\partial}{\partial \bar{P}} \left( \beta \frac{\Pi}{m'} + \frac{\partial}{\partial \Pi} \right) \right. \\
& \left. + \bar{\gamma}_{gf} \cdot \frac{\partial}{\partial \Pi} \left( \beta \frac{\bar{P}}{m} + \frac{\partial}{\partial \bar{P}} \right) + \gamma_{gg} \frac{\partial}{\partial \Pi} \left( \beta \frac{\Pi}{m'} + \frac{\partial}{\partial \Pi} \right) \right] W.
\end{aligned} \tag{4.3.19}$$

In this FP equation the COM and ST order parameters are coupled in three ways. The thermally-averaged forces  $\bar{f}^{th}$  and  $g^{th}$  depend on  $\bar{R}$  and  $\Phi$  (except for a virus in an otherwise uniform medium wherein there is only  $\Phi$ -dependence). Coupling is also provided through the  $\bar{R}$ ,  $\Phi$ -dependence of the friction coefficients. The cross-friction coefficients  $\bar{\gamma}_{fg}$  and  $\bar{\gamma}_{gf}$  provide additional coupling. However, for a virus in an isotropic medium,  $\bar{\gamma}_{fg}$  and  $\bar{\gamma}_{gf}$  vanish because the thermal average of any vector (e.g.  $(\bar{f}(0)g(t))^{th}$ ) is zero for an isotropic system.

To deduce some of the implications of (4.3.19) for viral STs, consider an isolated virus in an

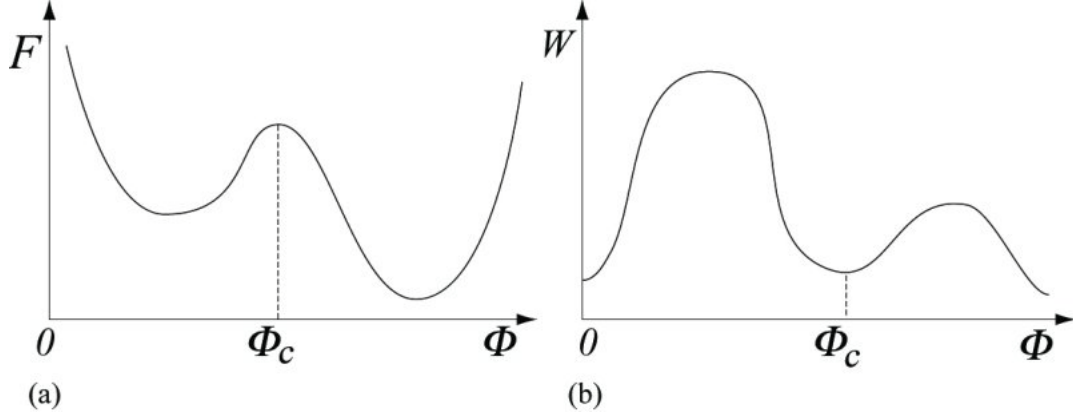
isotropic medium, and that the viral COM momentum is at equilibrium. In this case  $W = \tilde{W} \exp(-\beta P^2/2m)/Q_{CM}$  where  $Q_{CM}$  insures normalization for the COM part of  $W$ . In this case, the reduced order parameter distribution  $\tilde{W}(\Pi, \Phi, \tau)$  satisfies

$$\frac{\partial \tilde{W}}{\partial \tau} = - \left[ \frac{\Pi}{m'} \frac{\partial}{\partial \Phi} + g^{th} \frac{\partial}{\partial \Pi} \right] \tilde{W} + \varepsilon^2 \gamma_{gg} \frac{\partial}{\partial \Pi} \left( \beta \frac{\Pi}{m'} + \frac{\partial}{\partial \Pi} \right) \tilde{W}, \quad (4.3.20)$$

as can be verified upon substituting the product solution for  $W$  in (4.3.19). Let  $\ln Q = -\beta F$  for  $\Phi$ -dependent free energy  $F$ . Then  $g^{th} = -\partial F / \partial \Phi$  from which we see that (4.3.20) has the equilibrium solution that is proportional to  $\exp \left[ -\beta \left( \Pi^2 / 2m' + F \right) \right]$ . Note that  $\Pi$  ranges over all positive and negative values, while  $\Phi$  is strictly positive and  $F \rightarrow \infty$  as  $\Phi \rightarrow 0$  (as  $\Phi = 0$  represents a virus wherein all atoms are at the COM).

#### 4.4 Transition State Theory

In this section, TS theory for obtaining approximate solutions to the FP equation (4.3.20) in the inertial limit and estimating the free energy barrier of viral STs with their free energy profile sketched in Figure II.1 is presented as below. The TS ansatz is designed for the case where the barrier height is much greater than  $k_B T$ . Thus the system resides for an appreciable time in a given free energy well, and one may thus define a viral state as the subset of detailed slow variable configurations residing within the well. TS theory can be used to compute the evolving statistics of the subset of viruses whose state resides within a given well.



**Figure II.1** (a) a double well free energy profile  $F$  vs.  $\Phi$  implied by the hysteresis in NoV capsid ST and (b) the associated probability distribution  $W$  vs.  $\Phi$ .

As suggested in Figure II.1, the value of  $\Phi$  in the TS is  $\Phi_c$ , the location of the barrier separating the left ( $-$ ) and right ( $+$ ) wells. To start the analysis, define the total probabilities  $\tilde{W}_{\pm}$  of occupation of the ( $-$ ) and ( $+$ ) wells via

$$\tilde{W}_{-}(\tau) = \int_0^{\Phi_c} d\Phi \int d\Pi \tilde{W}(\Pi, \Phi, \tau), \quad \tilde{W}_{+}(\tau) = \int_{\Phi_c}^{\infty} d\Phi \int d\Pi \tilde{W}(\Pi, \Phi, \tau). \quad (4.4.1)$$

Integrating (4.3.20) over  $\Phi$  from  $\Phi_c$  to  $\infty$  (and similarly from  $0$  to  $\Phi_c$ ) and over all  $\Pi$  yields

$$\frac{d\tilde{W}_{\pm}}{d\tau} = \pm \int d\Pi \frac{\Pi}{m'} \tilde{W}(\Pi, \Phi_c, \tau). \quad (4.4.2)$$

In obtaining (4.4.2) we used the fact that because  $F \rightarrow \infty$  as  $\Phi \rightarrow 0$  or  $\infty$ , the probability of finding a virus in such extreme configurations is zero, i.e.  $\tilde{W}(\Pi, 0, \tau) = \tilde{W}(\Pi, \infty, \tau) = 0$ .

According to TS theory, in the right well  $\tilde{W}$  is essentially an equilibrium distribution and hence is proportional to  $\exp[-\beta(\Pi^2/2m' + F)]$ ; therefore the  $\tilde{W}$ -weighted average of  $\Pi$  is zero in the well, showing that there is no contribution to  $d\tilde{W}_{+}/d\tau$  except from a small

region of  $\Phi$  near  $\Phi_c$ . The TS ansatz for  $\tilde{W}(\Pi, \Phi, \tau)$  is developed as follows. When the barrier is much higher than  $k_B T$ , the two domains ( $\Phi$  to the left or right of  $\Phi_c$ ) act as if they were independent; hence  $\tilde{W}$  is essentially its equilibrium value for the left and right wells independently. Only infrequently is the barrier surmounted. Thus  $\tilde{W}$  is proportional to  $\exp[-\beta(\Pi^2/2m' + F)]/Z_-$  for  $\Phi < \Phi_c$  and  $\exp[-\beta(\Pi^2/2m' + F)]/Z_+$  for  $\Phi > \Phi_c$ , where

$$Z_- = \int_0^{\Phi_c} d\Phi \int d\Pi e^{-\beta(\Pi^2/2m' + F)}, \quad Z_+ = \int_{\Phi_c}^{\infty} d\Phi \int d\Pi e^{-\beta(\Pi^2/2m' + F)}. \quad (4.4.3)$$

With this and the definition of  $\tilde{W}_{\pm}$  in (4.4.1), the proportionality constants are determined and we obtain

$$\tilde{W} = \begin{cases} \exp[-\beta(\Pi^2/2m' + F)] \frac{\tilde{W}_-}{Z_-}, & \Phi < \Phi_c \\ \exp[-\beta(\Pi^2/2m' + F)] \frac{\tilde{W}_+}{Z_+}, & \Phi > \Phi_c. \end{cases} \quad (4.4.4)$$

However, near  $\Phi_c$  infrequent barrier-crossing takes place. The TS theory suggests that at  $\Phi_c$  the right-going viruses are predominantly coming from the left near-equilibrium well, and vice versa. Thus

$$\tilde{W}(\Pi, \Phi_c, \tau) = e^{-\beta(\Pi^2/2m' + F_c)} \left\{ \frac{\tilde{W}_-}{Z_-} \theta(\Pi) + \frac{\tilde{W}_+}{Z_+} \theta(-\Pi) \right\}, \quad (4.4.5)$$

$$\theta(x) = \begin{cases} 0, & x < 0 \\ 1, & x > 0, \end{cases} \quad (4.4.6)$$

and  $F_c$  is  $F$  evaluated at  $\Phi_c$ . With this ansatz, (4.4.2) becomes

$$\frac{d\tilde{W}_+}{d\tau} = \frac{1}{2}\omega \left[ e^{-\beta\Delta F_-^c} \frac{\tilde{W}_-}{Y_-} - e^{-\beta\Delta F_+^c} \frac{\tilde{W}_+}{Y_+} \right], \quad (4.4.7)$$

where

$$\omega = \left( \frac{2}{\pi\beta m'} \right)^{1/2}, \quad (4.4.8)$$

$$Y_- = \int_0^{\Phi_c} d\Phi e^{-\beta\Delta F_-}, \quad Y_+ = \int_{\Phi_c}^{\infty} d\Phi e^{-\beta\Delta F_+}, \quad (4.4.9)$$

$$\Delta F_{\pm} = F - F_{\pm}, \quad (4.4.10)$$

where  $F_-$  and  $F_+$  are  $F$  evaluated at the minimum of the left and right well respectively.

The superscript  $c$  on  $\Delta F_{\pm}^c$  indicates evaluation at  $\Phi_c$ . The quantity  $\omega$  is the thermal

average of  $|\Pi/m'|$  and hence has dimensions of  $\text{sec}^{-1}$ . As  $\tilde{W}_+ + \tilde{W}_- = 1$ , a second equation for

$\tilde{W}_-$  is not needed. Recalling that  $\tau = \varepsilon^2 t$ , one should add a factor  $\varepsilon^2 = m/m^*$  to the RHS of

(4.4.7) to get  $d\tilde{W}_+/dt$  ( $= \varepsilon^2 d\tilde{W}_+/d\tau$ ).

## Chapter 5 Molecular Dynamics/Order Parameter eXtrapolation (MD/OPX)

### 5.1 Characteristics of Order Parameters

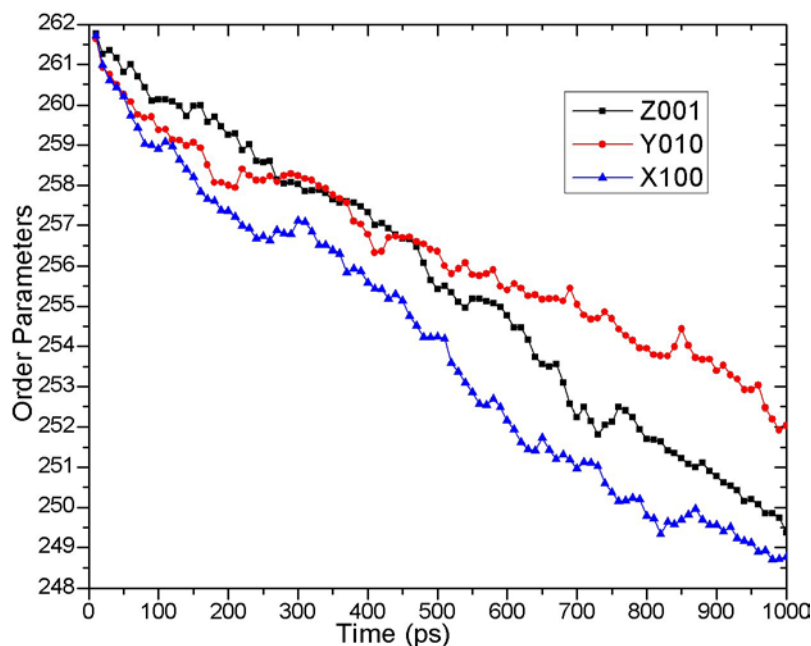
Newton's equations and statistical considerations can be used to argue that the order parameters  $\bar{\Psi}_{\underline{k}}$  constructed in Sect. 3.4 evolve slowly, i.e., on a timescale longer than the  $10^{-14}$ -second timescale of fast atomistic vibrations and collisions. As  $d\bar{s}_i/dt = \bar{p}_i/m_i$  for momentum  $\bar{p}_i$  of atom  $i$ , from Eqn. (3.4.6) one has

$$\begin{aligned} \frac{d\bar{\Psi}_{\underline{k}}}{dt} &= \frac{\bar{\Pi}_{\underline{k}}}{Nm}, \\ \bar{\Pi}_{\underline{k}} &\approx L^3 \sum_{i=1}^N U_{\underline{k}}(\bar{s}_i^0) \bar{p}_i. \end{aligned} \tag{5.1.1}$$

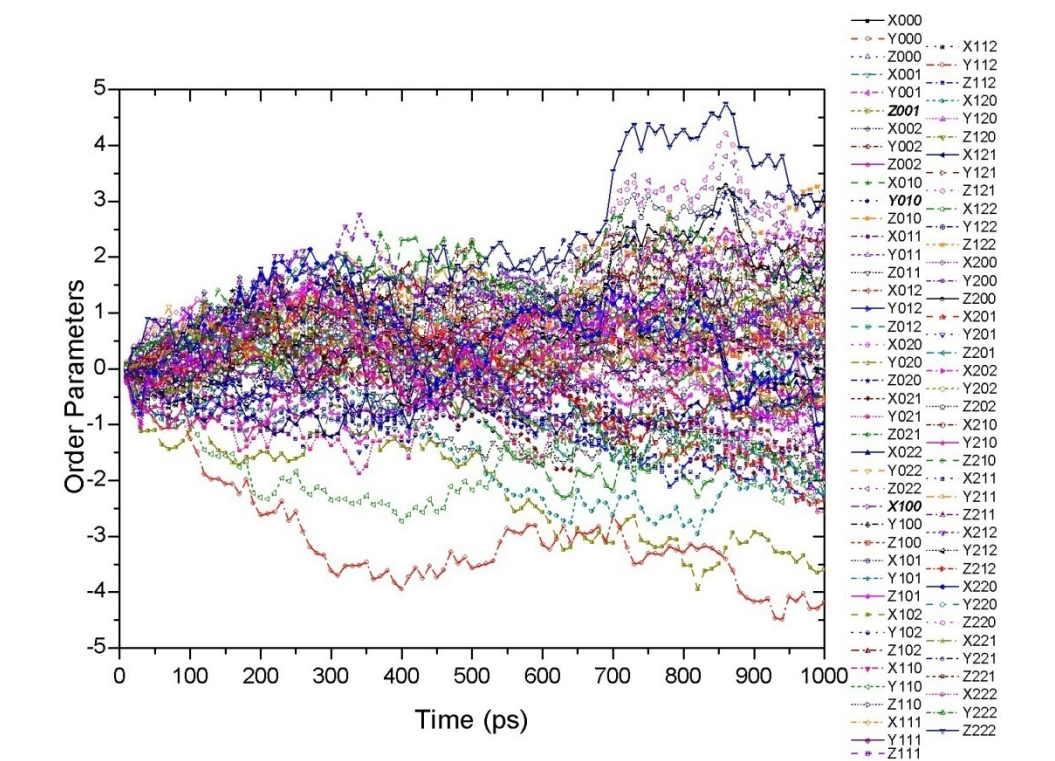
While  $\bar{\Psi}_{\underline{k}}$  has a sum of  $N$  atoms many of which have similar directions due to the smooth variation in  $U_{\underline{k}}$  with  $\bar{s}_i^0$ , the momenta have fluctuating direction and tend to cancel near equilibrium. Hence thermal average of  $\bar{\Pi}_{\underline{k}}$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

To demonstrate the slowly-varying characteristics of order parameters,  $3^3$  order parameters 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 (refer to Chapter 7 for simulation details) are plotted versus time. Figure II.2a shows time courses of the capsid order parameters 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 order parameters vary slowly and can

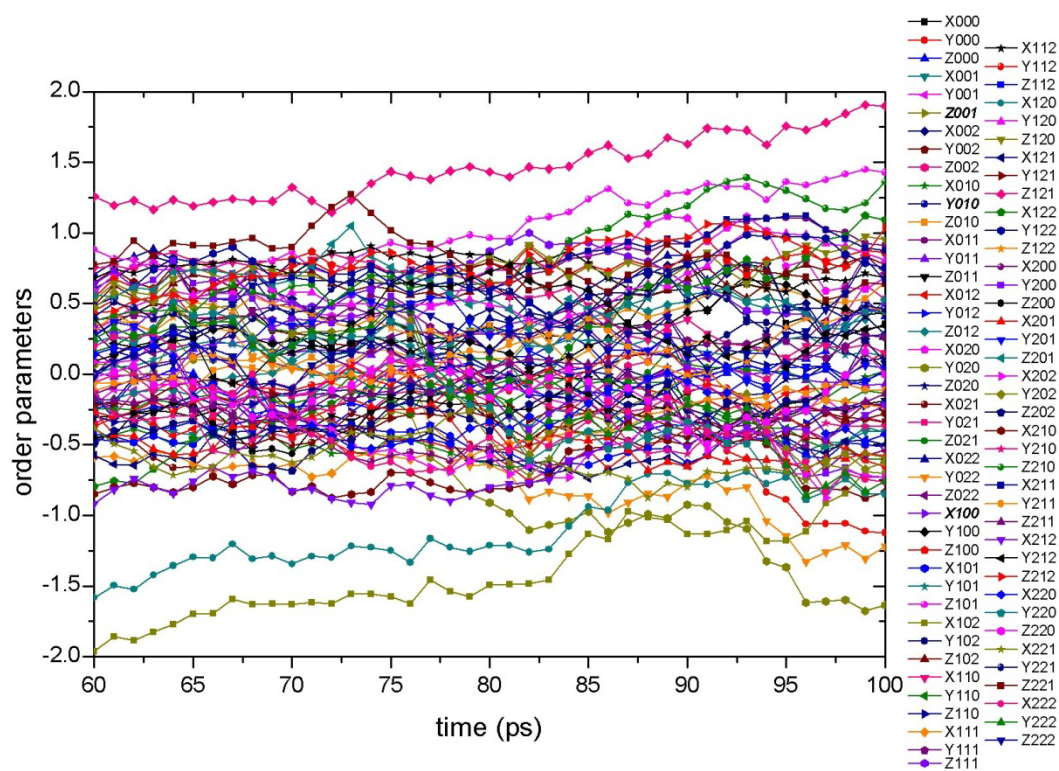
be readily extrapolated to long time on a timescale of ns. Other order parameters shown in Figure II.2b are seen to fluctuate rapidly around zero over 1ns. A closer look at these order parameters from 60ps to 100ps (Figure II.2c) shows that their values are comparatively stable on a timescale of 10ps, which implies their characteristic time periods are 10ps or longer. This means that all the order parameters 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 order parameters satisfy one of the criteria for the applicability of MD/OPX. The origin of the OP fluctuations in Figure II.2b 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 order parameters have longer characteristic time and can also be extrapolated over larger time intervals.



(a)



(b)



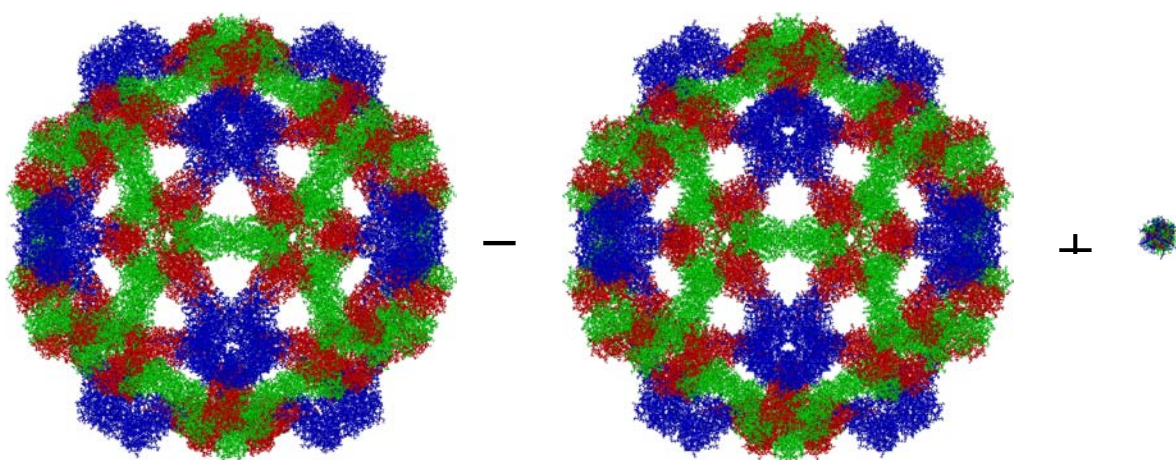
(c)

**Figure II.2** Time courses of CCMV capsid order parameters corresponding to Legendre



polynomials in the X, Y, and Z directions of the MD simulation: (a) three order parameters 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 can be readily extrapolated to long time on a timescale of ns, (b) other order parameters fluctuating rapidly around zero over 1ns and (c) a closer look at them from 60ps to 100ps showing that their characteristic time is roughly 10ps.

Another feature of order parameters is identified by investigating the coherent and residual contributions to the atomic configuration along its simulated trajectory. Figure II.3 shows a snapshot of the CCMV capsid after 1ps MD 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  $3^3$  order parameters are found to reflect overall capsid structure with the residuals close to zero. This suggests that the slow collective motions in viral STs are captured with coherent coordinates computed through the slowly-varying order parameters while the fast motions including atomic vibrations and collisions are accounted for by the rapidly-fluctuating residuals.



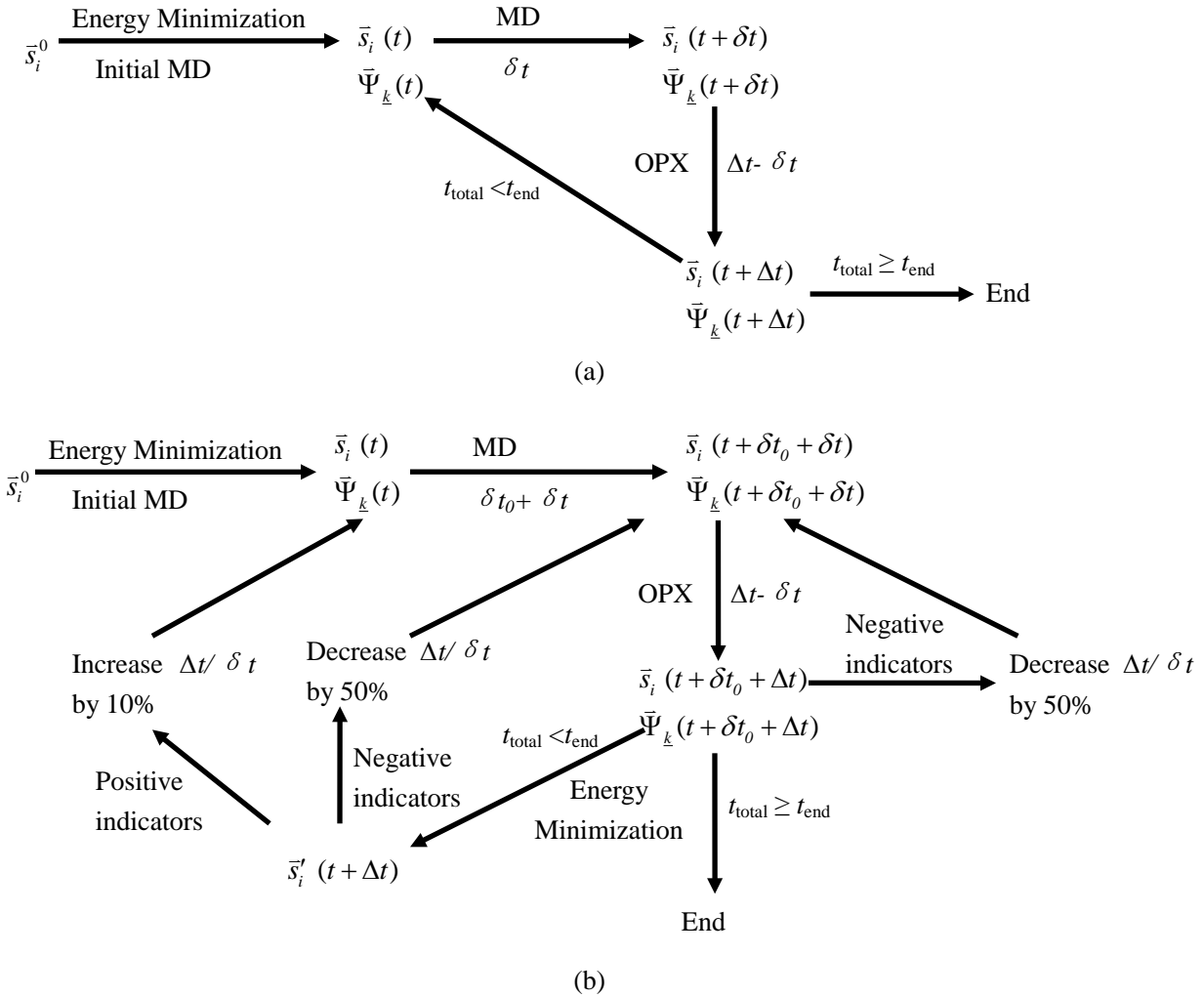
**Figure II.3** A snapshot of the CCMV capsid after 1ps MD simulation showing that the atomic coordinates are equal to their coherent contribution plus small residuals.

## 5.2 MD/OPX Approach

In the above section, the order parameters automatically constructed with orthogonal basis functions of atomic variables are shown to evolve on timescales much larger than the  $10^{-14}$ -second fast atomic vibrations and collisions based on both statistical analysis of Newton's equations and calculations on output atomic trajectories of viral molecular dynamics simulations. AMA of dynamical nanosystems [11, 102] demonstrates the existence of a stochastic (Fokker-Planck or Smoluchowski) equation and equivalent Langevin equations for the dynamics of order parameters  $\bar{\Psi}_{\underline{k}}$ . In this section, we will present MD/OPX, a computational approach to solving the Langevin equations of order parameters 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 order parameters via Eqn. (3.4.5). In a small time interval  $\delta t$  (e.g., 100 MD timesteps), an MD code output can be used to generate  $\bar{\Psi}_{\underline{k}}(t + \delta t)$ . Then, the change  $\bar{\Psi}_{\underline{k}}(t + \delta t) - \bar{\Psi}_{\underline{k}}(t)$  divided by  $\delta t$  represents the deterministic part of the Langevin dynamics. With this, one approximates  $\bar{\Psi}_{\underline{k}}(t + \Delta t)$  for a timestep  $\Delta t$  via  $\bar{\Psi}_{\underline{k}}(t + \Delta t) \approx \bar{\Psi}_{\underline{k}}(t) + \Delta t [\bar{\Psi}_{\underline{k}}(t + \delta t) - \bar{\Psi}_{\underline{k}}(t)] / \delta t$ , where  $\Delta t \gg \delta t$  but smaller than the characteristic time for OP dynamics. In this way, one may advance  $\bar{\Psi}_{\underline{k}}$  via a sequence of dual  $(\delta t, \Delta t)$  cycles. We use Eqn. (3.4.4) to reconstruct the all-atom configuration after the OP advancement to  $t + \Delta t$ , taking the residual  $\bar{\sigma}_i$  to be its value at  $t + \delta t$ . We justify the use of  $\bar{\sigma}_i(t + \delta t)$  to approximate  $\bar{\sigma}_i(t + \Delta t)$  because we find that (1) with a sufficient number of basis

functions  $U_k$  chosen the  $\bar{\sigma}_i$  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. The flowchart in Figure II.4a shows the above dual timestep MD/OPX algorithm. Starting from a PDB structure, an initial minimization and an initial MD run are applied before MD/OPX cycling commences.



**Figure II.4** Schematic flowchart of MD/OPX (a) and its implementation with dynamic adaptive run parameter adjustment (b): energy minimization and a short  $\delta t_0$  MD run are applied to anneal the simulated all-atom structure after OPX and run parameters (notably  $\delta t$  and  $\Delta t$ ) are adjusted dynamically according to structural and dynamic indicators to optimize the balance between program accuracy and CPU speed.

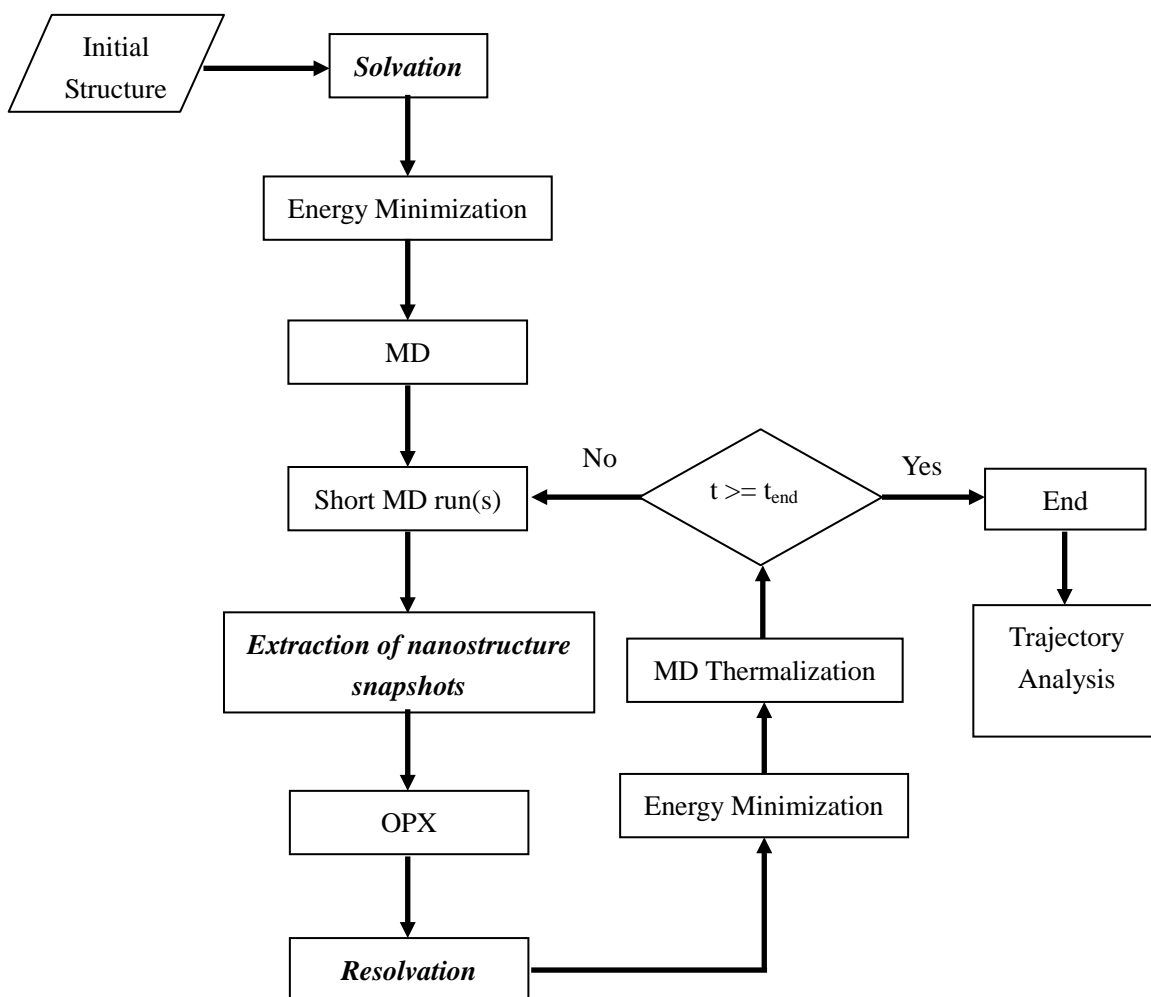
A variety of steps were taken to optimize and stabilize the MD/OPX simulations. First, the OPX over  $\Delta t$  deforms all space continuously. This can lead to unphysical atomic structures, such as stretched bonds, unreasonable bond angles and dihedrals, close/overlapping atoms, etc. Thus we perform an energy minimization and a short MD run to anneal these unphysical structures after obtaining the all-atom structure that is generated from extrapolated order parameters. Next we implement an adaptive cycling. Simulation parameters including  $\delta t$ ,  $\Delta t$ , the amount of energy minimization steps and the length of the short MD run ( $\delta t_0$ ) 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 increases  $\Delta t/\delta t$  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 for adaptive MD/OPX is shown in Figure II.4b.

Several cautionary notes are in order. The deterministic forces of the Langevin equations

contribute to  $\bar{\Pi}_{\underline{k}}$  (associated momenta of  $\bar{\Psi}_{\underline{k}}$ ) to order  $\Delta t$ , 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  $\delta t$  and  $\Delta t$ . Another source of error is that  $\delta t$  might not be large enough. If  $\delta t$  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 [11]. We address this by performing simulations to determine an acceptable  $\delta t$ , for which results become insensitive to further  $\delta t$  increases. Thus  $\delta t$  cannot be too short, but the extrapolation scheme requires that  $\delta t$  also be short relative to the characteristic time for which the order parameters change appreciably. Thus, one must make use of the adaptive choice of  $\delta t$  and  $\Delta t$  provided in MD/OPX to achieve an acceptable balance between accuracy and efficiency.

### 5.3 MD/OPX Implementation and Optimization

While Figure II.4 shows the flowchart of MD/OPX for simulating nanostructures in vacuum (i.e., gas phase) [12, 13], modules have been developed to account for water molecules and ions for simulating nanostructures solvated in host media. Figure II.5 shows the workflow of MD/OPX, which is presently implemented by using NAMD[75] as the MD platform. VMD is used for “solvation” and “resolution” of simulated nanostructures and analysis of output trajectories.



**Figure II.5** Schematic flowchart of MD/OPX implemented by using NAMD and VMD as the platform for simulating nanostructures in vacuum or solvated in host media. Modules highlighted in bold are needed for simulations of nanostructures in host media.

MD/OPX starts with an initial structure that can be determined from X-ray crystallography, NMR spectroscopy, and cryoEM. For system preparation, the structure is solvated in water and neutralized under certain pH and ionic strength if needed by using VMD. The resulting system along with its structural topology is then taken as input to NAMD for initial energy minimization and equilibration, after which the simulation is performed in MD/OPX cycles until it reaches the end of specified time. In each MD/OPX cycle, short MD run(s) are used to generate the time evolution of atomic configuration of the system. With several snapshots

chosen from the system trajectory, nanostructure without water molecules and ions is extracted for calculating the structural order parameters (Sect. 3.4) and therefore estimating their rates of changes, i.e., time derivatives, which are then used to extrapolate order parameters over long time. The nanostructure atomic configuration is then reconstructed from newly obtained order parameters. A Fortran program was written to read NAMD output all-atom structures, calculate the resultant order parameters by solving Eqn. (3.4.5), extrapolate the order parameters for  $\Delta t$ , generate an atomic configuration at  $t + \Delta t$  with the extrapolated order parameters. The resulting all-atom structure is resolvated in the host medium that keeps the same number of water molecules and ions without change in the system structural topology. Energy minimization and MD thermalization are applied to the resulting system before running short MD for another cycle.

In the “resolution” module, the OPX resulting nanostructure is first put back to the host medium by using the final configuration of one short MD run, which keeps the same number of water molecules and ions (i.e., same system structural topology) as needed for continuing the simulation in NAMD. Water molecules and ions that are within a certain distance of the nanostructure (denoted  $d_s$ ) are removed and then redistributed as follows since many of them may overlap with the nanostructure or be embedded in it. First, a water box that is large enough to solvate the nanostructure with at least a water boundary of  $d_s$  ( $340^3 \text{ \AA}^3$  for simulation of CCMV capsid swelling) is prepared and equilibrated under 298.15K and 1atm by running MD for 100ps. With the nanostructure placed in the center of the equilibrated water box, the coordinates of water molecules, which are selected as within  $d_s$  and not overlapping with the

nanostructure, are retrieved to replace those of the above removed water molecules and ions. Note that the number of the retrieved water molecules may not be exactly the same as that of the needed for redistribution due to the water density difference between the two systems. Thus  $d_s$  is adjusted iteratively by small increments for selecting water molecules in the equilibrated water box until they are enough for the redistribution of the removed water molecules and ions. The ions are moved to positions that are nearest to their original ones and the overlapping water molecules are removed. This generates the final configuration that has no overlapping between the nanostructure and host particles, and maintains local density and pressure of the host medium.

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  $\delta t$  and  $\Delta t$ , (2) different types of basis functions  $U_{\underline{k}}$ , (3) automated choice of the number of order parameters, and (4) compute  $\bar{\Psi}_{\underline{k}}$  at the beginning and end of the  $\delta t$  MD interval and use them for simple Euler extrapolation, or compute  $\bar{\Psi}_{\underline{k}}$  multiple times during the  $\delta t$  MD interval and then smooth the  $\bar{\Psi}_{\underline{k}}$  profiles via least squares fitting to better estimate  $d\bar{\Psi}_{\underline{k}}/dt$  for carrying out the extrapolation to  $t + \Delta t$ .

MD/OPX has also been optimized by (1) averaging a set of short replica MD runs with random atomic velocity initializations to compute ensemble-average rates of change in order parameters, which allows larger timestep for extrapolation, (2) updating the reference configuration for constructing system order parameters at a certain frequency such that the latest collective motions of the structure can be effectively captured, (3) improving the maturing



of OPX resulting structure by applying energy minimization, graduate heating, and short-time MD equilibration before running short replica MD bursts for OPX, and (4) choosing optimal OP timestep for extrapolation and atomic acceleration indicators to evaluate the OPX resulting structure to achieve adaptive variation of MD/OPX parameters.

## Part III. Results and Discussion

### Overview

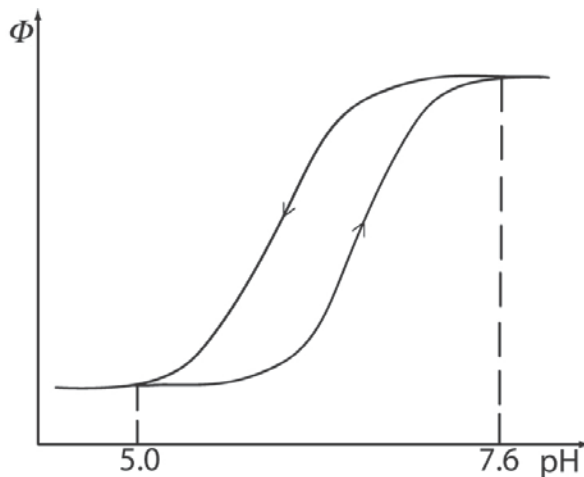
In this part, methods developed in Part II are used to study viral structural transitions, notably in NϕV and CCMV. In Chapter 6, the transition state theory derived from the multiscale viral dilatational structural transition model is applied to NϕV capsid, for which the free energy barrier of its structural transition from the capsid immature form (procapsid) to the mature capsid is estimated roughly [9]. In Chapter 7, MD/OPX is demonstrated on CCMV capsid with a 1ns simulation of the capsid swollen state analyzed quantitatively and the results are shown to agree well with those of a direct MD simulation[12]. After the demonstration, MD/OPX is applied to the shrinkage of swollen CCMV capsid in vacuum in Chapter 8. 200ns MD/OPX simulation of the swollen state of CCMV capsid reveals that it undergoes significant energy-driven shrinkage, which is a symmetry-breaking process involving local initiation and front propagation [13]. In Chapter 9, with modules added to account for water molecules and ions, MD/OPX is applied to the swelling of native CCMV capsid solvated in host media [14]. Simulation results show that the N-terminal arms of capsid proteins undergo fast fluctuations during early stage of the simulation and their structural changes trigger global expansion of the capsid. Swelling of CCMV capsid is also a symmetry-breaking process involving local initiation and front propagation. Prospects for using our simulation predictions to guide experiments are discussed.

## **Chapter 6 Application of Transition State Theory to N $\omega$ V Capsid**

### **6.1 Hysteresis, Irreversibility and the Free Energy Profile in Viral Structural Transitions**

Hysteresis of capsid assembly/disassembly in hepatitis B virus (HBV) has been observed and studied with a chemical kinetic model [125]. CCMV capsid undergoes a pH and metal ion dependent reversible swelling transition between close native and open swollen forms [1, 15, 18, 39, 40, 45]. This expansion is reversible and titration experiments revealed that it involves hysteresis at low salinity ( $I=0.2M$ ) [126]. Hysteretic effects have also been found in the maturation of N $\omega$ V [41]. N $\omega$ V undergoes large conformational changes from a procapsid form (480 Å in diameter) to a compact capsid form (410 Å in diameter) when pH is decreased from 7.6 to 5.0. The transition takes less than 100 ms and is accompanied by a slow autoproteolysis (taking hours) corresponding to the cleavage of 70 kDa coat proteins to 62 kDa and 8 kDa proteins. The conformational rearrangement is initially reversible until about 15% of the cleavage events are completed, at which point the particles are locked into the capsid conformation, regardless of pH [10, 47, 48]. A further study on a cleavage-defective mutant (N570T) of N $\omega$ V showed that the transition from procapsid to capsid in the mutant is reversible, and that the reverse process is much slower, with some capsids never reexpanded after 4 days of dialysis against pH 7.6 buffer, and the reexpanded procapsids display slightly different properties than the original capsid, suggesting hysteresis but not complete reversibility in the transition [41]. Irreversible poliovirus conformational changes are found in receptor-mediated cell entry, during which its coat protein VP4 and the N terminus of VP1 are externalized. Two

putative cell entry intermediates (135S and 80S particles) are formed and they are both about 4% larger than the native virions (160S particles) [7, 8, 43, 44, 46]. The viral receptor behaves as a classic TS theory catalyst, facilitating the ST from native virions to 135S intermediate particles by lowering the activation energy for the process by 50 kcal/mol [46].



**Figure III.1** The cleavage-defective mutant of N $\omega$ V undergoes hysteretic ST when pH in the host medium changes. The curve indicates values of  $\Phi$  for which the virus/host system's free energy is a minimum.

Hysteresis in the N $\omega$ V ST implies that there are two structurally-distinct states of the virus for a given range of host pH. Such states correspond to free energy wells, suggesting a free energy landscape with double-well character. Viruses make transitions between fluctuating states defined by deep, local free energy wells. The residence time within a given well is determined by the intensity of thermal fluctuations, the mass and size of the virus, the number of states within the well and the height of free energy barriers that must be crossed in exiting the well.

As suggested in Figure III.1, when pH changes in the host medium, an order parameter  $\Phi$  (i.e. the relative overall size of the virus as defined in Sect. 4.1) of the cleavage-defective

mutant of N $\omega$ V can make a hysteric transition between two states with distinct structures, i.e. residing within different free-energy wells. Underlying the transition is a double-well free energy profile as suggested in Figure II.1. In more complex systems, e.g. for HK97 bacteriophage [49, 50], there are multiple wells in a multidimensional order parameter space and transitions between them. Even for systems supporting a single transition, there are likely several key order parameters so that a true picture of the ST must follow from an analysis of landscapes in higher dimensional order parameter space. A single order parameter model supporting two states with distinct structures implied by different ranges of  $\Phi$  is suggested in Figure III.1 along with the associated probability distribution.

## 6.2 Free Energy Barrier of N $\omega$ V capsid Structural Transition

Time-resolved small-angle X-ray scattering (TR-SAXS) experiments show that the capsid ST during N $\omega$ V maturation occurs on timescales in the range of 0.1 to 100 sec with the detection of a fast-forming intermediate in the transition [10]. To evaluate this in our formulation via TS theory, consider the purely forward contribution to (II.5.7) as recast in the form

$$\frac{d\tilde{W}_+}{dt} = \frac{1}{2Y_-} \varepsilon^2 \omega \exp\left[-\frac{\Delta F_-^c}{k_B T}\right] (1 - \tilde{W}_+). \quad (6.2.1)$$

Several factors are seen from (6.2.1) to limit the rate of a viral ST: (1) a factor 1/2 as only viruses undergoing purely forward transitions at the TS contribute, (2) the  $\varepsilon^2$  factor expressing the great size of a virus relative to a single atom (i.e.  $\varepsilon^2 \approx 10^{-6}$ ), (3)

$\exp[-\Delta F_-^c / k_B T]$  expressing a free energy barrier effect, and (4)  $Y_-$  introducing entropic effects that account for the number of  $\Phi$  configurations collectively labeled the  $(-)$  state in the left free energy well.

The rate law (6.2.1) is in the form  $d\tilde{W}_+/dt = k(1 - \tilde{W}_+)$ . From the aforementioned experimental data, the rate coefficient  $k$  varies from 10 to 0.01  $\text{sec}^{-1}$ . To use this data in estimating  $\Delta F_-^c$ , we adopt the following assumptions.  $Y_-$  is the range of  $\Phi$  in the  $(-)$  well for which  $\Delta F_-^c > k_B T$ . As a crude estimate, we use the fact that a change in the radius upon transition is about 16% for N $\omega$ V capsid and then take  $Y_-$  to be smaller than this, in particular 1% of the capsid radius, i.e.  $Y_- \approx 0.01$ . Using the carbon atom mass as typical for the average atom in the capsid, a capsid radius of 198 Å and  $\varepsilon^2 = m/m^* \approx 10^{-6}$ ,  $m'$  is given by (II.1.7), i.e.

$$m' \equiv \varepsilon^4 \sum_{i=1}^N m_i s_i^{02} \Theta_i \approx \varepsilon^2 \times \frac{12.01 \times 10^{-3} \text{ kg} \cdot \text{mol}^{-1}}{6.022 \times 10^{23} \text{ mol}^{-1}} \times (198 \times 10^{-10} \text{ m})^2 = 7.73 \times 10^{-48} \text{ kg} \cdot \text{m}^2. \quad (6.2.2)$$

This yields

$$\omega = \left( \frac{2}{\pi \beta m'} \right)^{1/2} \approx \left( \frac{2 \times 1.38 \times 10^{-23} \text{ m}^2 \cdot \text{kg} \cdot \text{s}^{-2} \cdot \text{K}^{-1} \times 298.15 \text{ K}}{\pi \times 7.73 \times 10^{-48} \text{ kg} \cdot \text{m}^2} \right)^{1/2} = 1.84 \times 10^{13} \text{ s}^{-1}. \quad (6.2.3)$$

Using this and (6.2.1) to evaluate  $\Delta F_-^c$  in kcal/mol, we have

$$\Delta F_-^c = -\frac{N_A}{4186 \text{ J} \cdot \text{kcal}^{-1}} k_B T \ln \frac{2kY_-}{\varepsilon^2 \omega} \approx \begin{cases} 10.86 \text{ kcal/mol}, k = 10 \text{ s}^{-1} \\ 14.95 \text{ kcal/mol}, k = 0.01 \text{ s}^{-1} \end{cases}. \quad (6.2.4)$$

The transition of native poliovirus (160S particle) to an infectious intermediate (135S particle) during cell entry is determined to have a free energy barrier of 30 kcal/mol without the cell receptor [46, 127]. Our calculations based on crude estimates show that the N $\omega$ V capsid ST

has a lower free energy barrier and it depends on how we change the pH to induce the ST. When pH is lowered from 7.6 to 5.0, it occurs on 0.1 sec timescale with a free energy barrier of 10.86 kcal/mol; with a less drastic pH change (pH lowered to 5.8 or 5.5 instead of 5.0), the ST occurs more slowly on a 100 sec timescale with a higher free energy barrier of 14.95 kcal/mol. The activation enthalpy for the poliovirus ST is known to be lowered by 50 kcal/mol in the presence of cell receptor and capsid-binding drugs for poliovirus are shown to inhibit the receptor-mediated ST through a combination of enthalpic and entropic effects. We suggest that viral STs must be understood in terms of free energy barriers and the entropics of viral states, as well as inertial effects (manifest in the  $\varepsilon^2$  factor in (6.2.1)). Inertial effects emerge from the large contrast in the mass of a whole virus, or one of its structural units, and that of an atom. This is one way in which the timescale of viral migration or overall structural change is much longer than that of individual atomic vibrations and collisions. However, there is also the effect of energy barrier crossing from one free energy well to another which implies that a virus resides in a given well for an extended time, only occasionally crossing the barrier separating wells where the virus experiences an unlikely thermal fluctuation. Finally entropics can also inhibit the rate of exiting a free energy well, i.e. a virus spends an extended period of time exploring many detailed configurations before making a ST. For example, this is why capsid assembly is much slower than disassembly.

The TS estimate of the free energy barrier of NoV capsid ST provides an upper bound, i.e. the single order parameter model likely misses lower energy barrier pathways for a viral capsid transition. However, it can also be argued that an overall dilatational fluctuation could provide

more room for capsid structural units to translate and rotate relative to their neighbor units and thereby initiate a ST. In either case, the single order parameter model of Chapter 4 can be used as a baseline to interpret experimental results in terms of free energy barrier and entropic effects that can then be compared with estimates from more complete models.

The AMA approach enables one to analyze viral STs from first principles and a calibrated interatomic force field. Our formulation can be generalized for multiple order parameter models to account for lower free energy barrier pathways for viral STs. The theory with its all-atom description can be applied to nonviral nanoparticles as well.



## **Chapter 7 MD/OPX Validation with CCMV Capsid Structural Transitions**

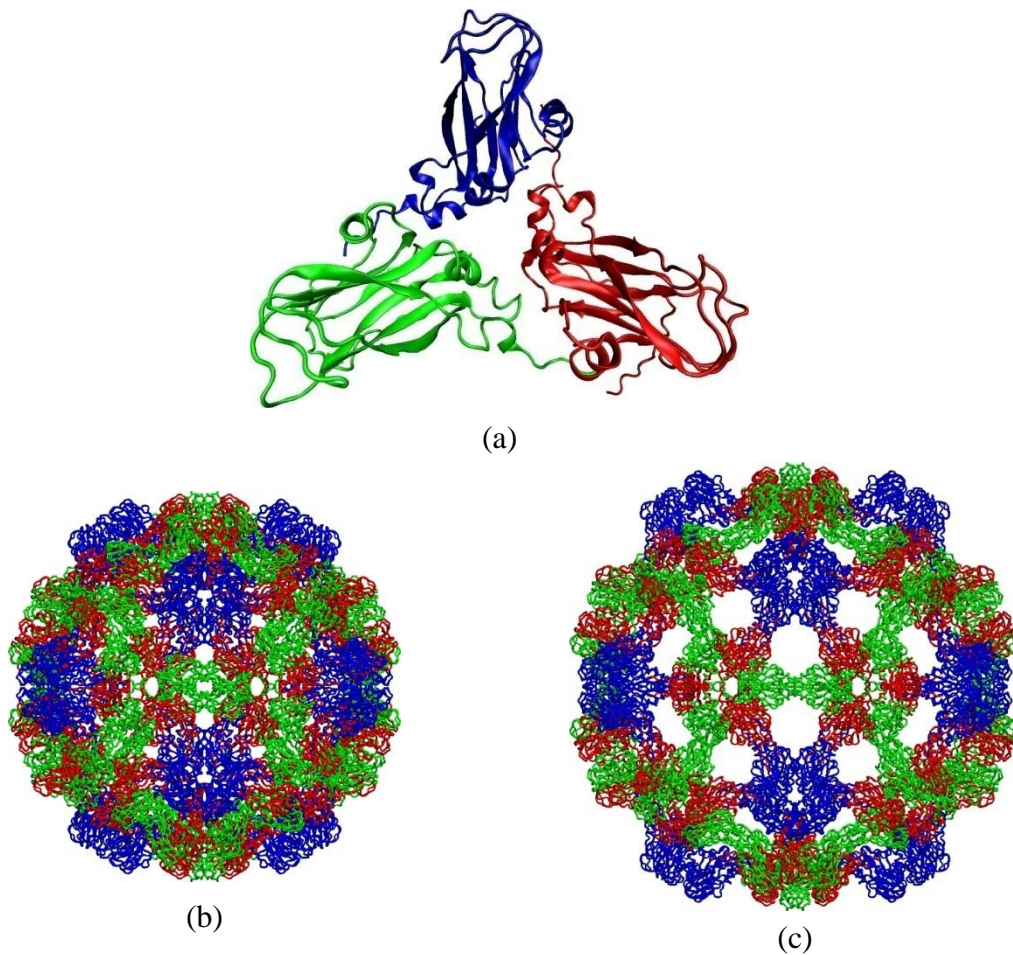
### **7.1 Background of CCMV and its Capsid Structural Transitions**

As reviewed in Chapter 1, 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[18]. 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. As such, one asymmetric unit (i.e., protomer) of the capsid includes three subunits that are colored in blue for A, red for B, and green for C in Figure III.2. The icosahedral capsid can also be divided into 12 pentamers and 20 hexamers with 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)[1, 18]. 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.[1], the swollen capsid (Figure III.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.



**Figure III.2** (a) Crystal structure of a CCMV protomer with its three quasi-threefold related subunits colored in blue (A), red (B) and green (C). (b) Native CCMV capsid organized in 12 pentameric and 20 hexameric capsomers with 5 blue A subunits in each pentamer and 3 red B and 3 green C subunits in each hexamer. (c) The swollen CCMV capsid generated

computationally by rigid-body translations and rotations of the pentamers and hexamers according to the expansion scheme proposed by Liu et al.[1].

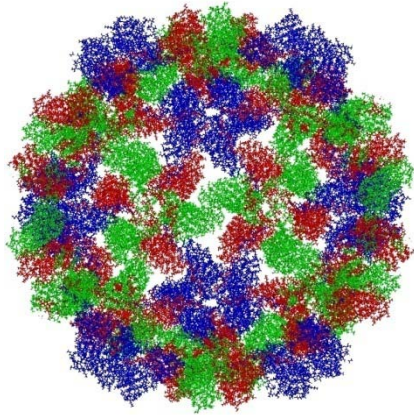
## **7.2 Order Parameters Capturing Capsomer Dynamics during CCMV Capsid Expansion**

To demonstrate the viability of our order parameters, 4 intermediate structures and the final swollen one were generated from native CCMV capsid equally distributed along its expansion path as described below. Order parameters 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 order parameters 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.

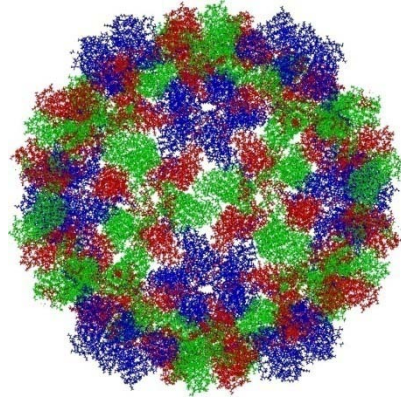
Starting from the native CCMV capsid, 5 steps were applied to transform the native capsid structure into the swollen structure with 4.8Å-translation and 1.8°-rotation for pentamers and 4.25Å-translation and 1.6°-rotation for hexamers in each step. The inner shell view of the 5 result structures are shown in Figure III.3 (a1) to (a5)). With these 5 atomic structures, we calculated the order parameters 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 order parameters were compared with the 5 original structures to investigate the performance of the order parameters.

Four sets of calculations were performed to evaluate the effects of varying the number of order parameters, notably  $3^3$ ,  $4^3$ ,  $5^3$  and  $6^3$ . For  $n^3$  order parameters, Legendre polynomials of

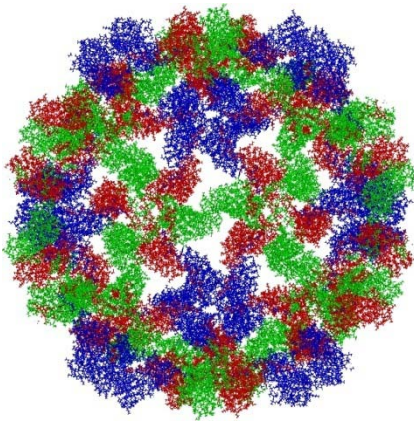
order  $(0,1,\dots,n-1)$  in the X, Y and Z directions were used (see Sect. 3.4). Figure III.3 (b1) to (b5) show a sample set of structures visualized with the coherent contribution of atomic coordinates calculated from  $3^3$  order parameters. They are found to reflect the overall expansion of the capsid, although the openings formed around the quasi-threefold axes during the expansion are lost. As the number of order parameters is increased from  $3^3$  to  $6^3$ , the calculated coherent structures become closer to their original atomic structures.



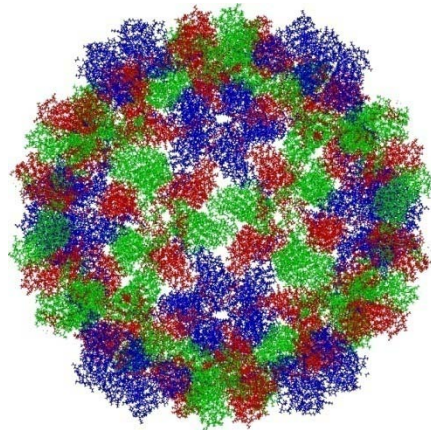
(a1)



(b1)

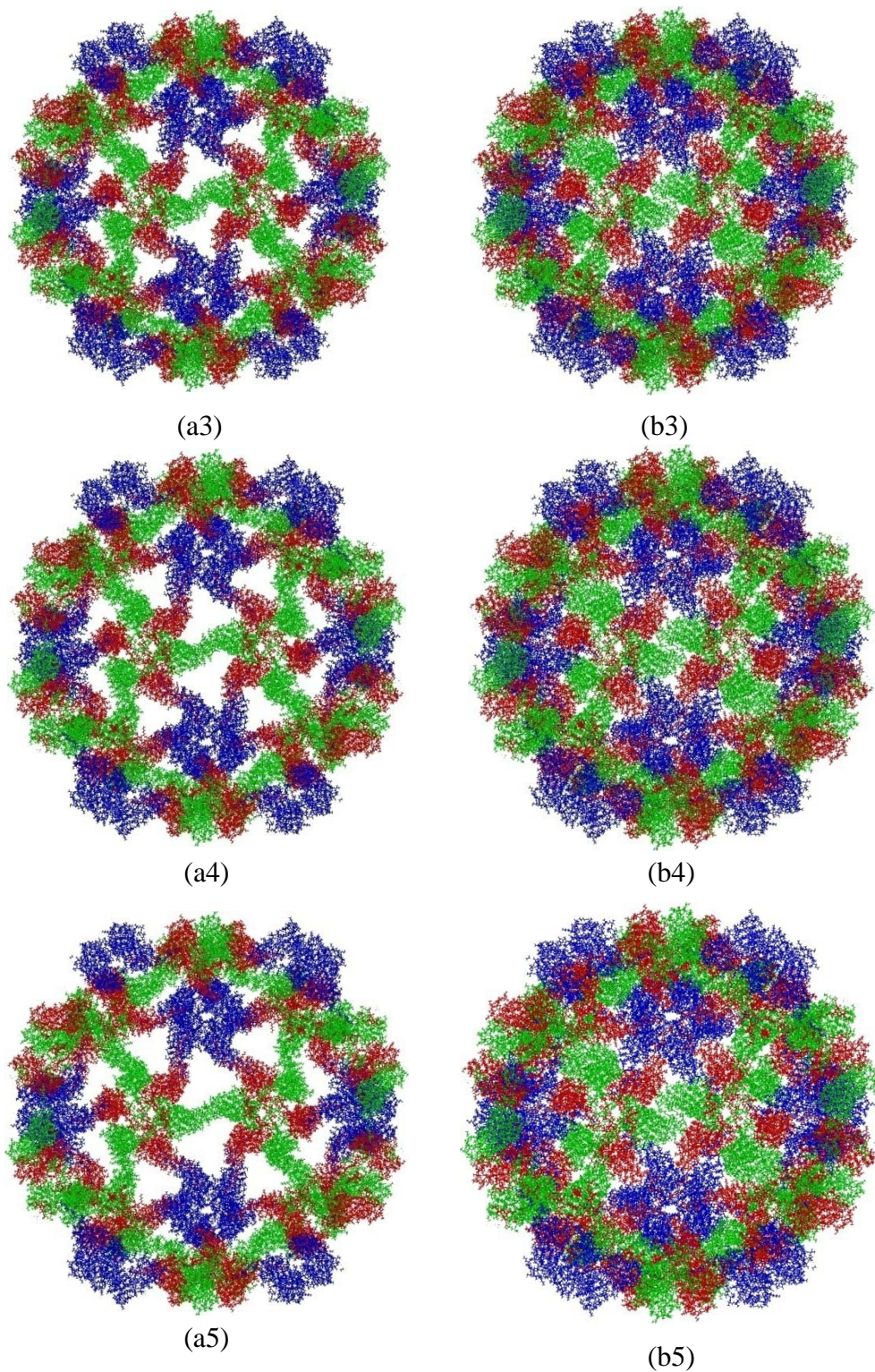


(a2)



(b2)

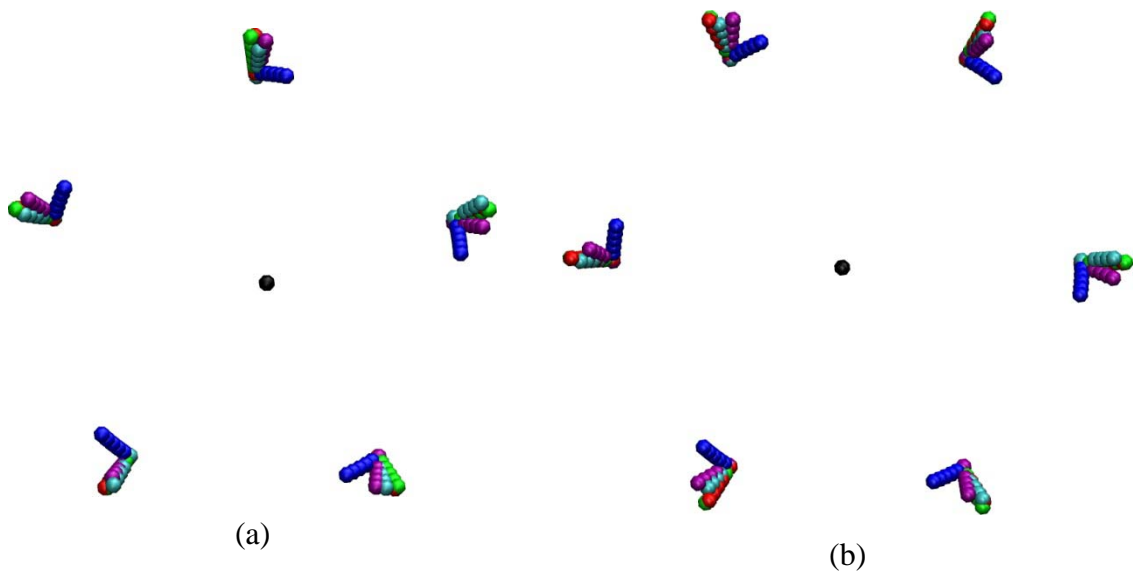




**Figure III.3** (a1) to (a5) in the left column show the inner shell view of the 4 intermediate CCMV capsid structures and the final swollen one generated from the native capsid and (b1) to (b5) in the right column show their corresponding coherent structures calculated from  $3^3$  order parameters.

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  $6^3$  order parameters. The center of mass (CM) positions of the protein subunits included in the pentamer from these structures were visualized in Figure III.4a, and similarly for the hexamer in Figure III.4b. In the figure, the black bead represents the COM position of the capsid, which stays at the origin for all structures; blue beads represent COM 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  $6^3$  order parameters. Top view of the blue beads slightly off the symmetry axes of the pentamer and hexamer in Figure III.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  $3^3$  order parameters (red beads) and the accuracy increases with the number of order parameters, and partial of the pentamer and hexamer rotation is captured with  $5^3$  (cyan beads) and  $6^3$  (purple beads) order parameters while it is not revealed with  $3^3$  (red beads) or  $4^3$  (green beads) order parameters. 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 order parameters 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 order parameters exceeds the translation DoF, the capsomer rotation starts to be captured. This suggests that our order parameters capture the ST dynamics when enough order parameters are used. The positions and orientations of beads in a coarse-grained model are in a sense a subset of our order parameters. However, as we construct our order parameters 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.



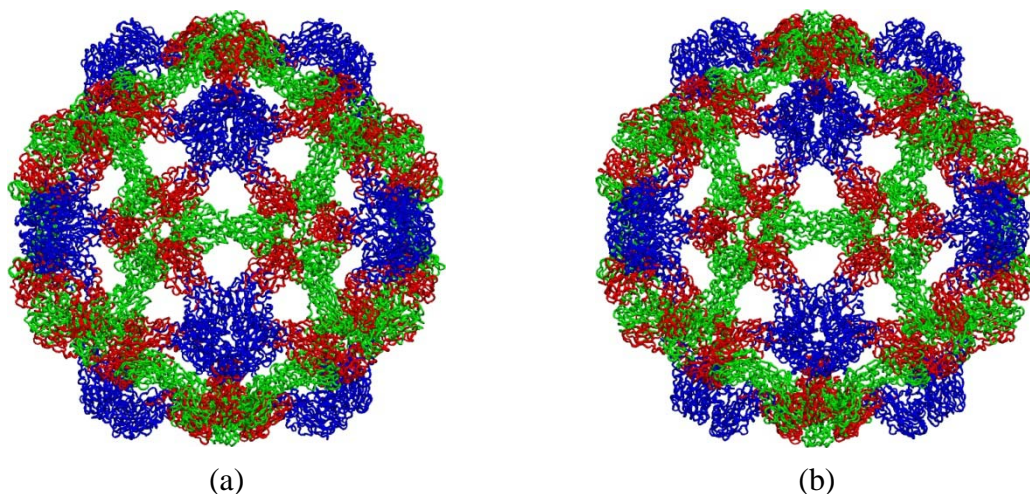
**Figure III.4** Demonstration of using order parameters to capture the translation and rotation of a pentamer (a) and a hexamer (b) during the native CCMV capsid expansion: the black bead represents the COM position of the capsid, which stays at the origin for all structures; blue

beads represent COM positions of the protein subunits in the chosen pentamer and hexamer from the 5 original capsid structures; while the red, green, cyan and purple beads correspond to the protein subunits from the 5 coherent capsid structures calculated from  $3^3$ ,  $4^3$ ,  $5^3$  and  $6^3$  order parameters.

### 7.3 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 (see Sect. 8.1). To study the evolution of our order parameters of Sect. 3.4 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  $3^3$  order parameters for capturing capsid shrinkage dynamics. 1000 MD/OPX cycles were run with each cycle composed of 100 1fs-MD steps, i.e.  $\delta t=100\text{fs}$  and one OPX of 900 equivalent MD steps, i.e.  $\Delta t=1000\text{fs}$ . Results, including snapshots of the simulated trajectories, time courses of the capsid radii, root mean square deviation (RMSD) from the starting structure, individual order parameters and program timing are compared with those of a direct MD simulation as follows.



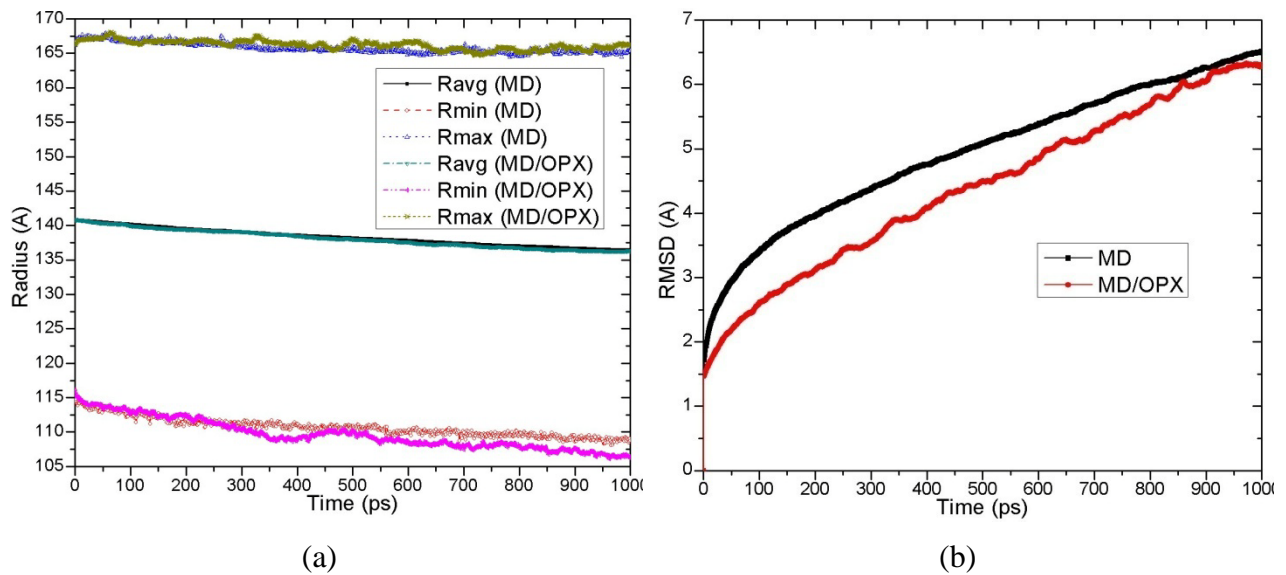


**Figure III.5** Ribbon representations of the output structures of the swollen CCMV capsid after 1ns simulation using MD (a) and MD/OPX (b).

The structures after 1ns simulation using MD (a) and MD/OPX (b) are shown in Figure III.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 Figure III.6a. 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. Figure III.6b 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  $3^3$  order parameters were used, the agreement was excellent. This shows that while the order parameters may not capture all the details, the residuals (see Eqn. II.4) retain the details lost by the order parameters. This suggests that completeness of the set of order parameters 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).



**Figure III.6** Comparison of 1ns MD and MD/OPX simulations: (a) Time courses of the

average, minimum and maximum radii of the CCMV capsid backbone (**b**) RMSD of atomic positions between the simulated CCMV capsid structure along the trajectory and its starting structure.

## 7.4 MD/OPX Performance

Above simulations 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 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 increasing the OP timestep. By using ten replica short MD runs with random atomic velocity initializations for OPX, the average OP timestep obtained from adaptive MD/OPX cycles can be largely increased to 60ps. However, additional energy minimization and short MD thermalization are needed to anneal the OPX resulting structure, which could downgrade the program running performance (see Sects. 5.3 and 8.2). 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. Also as the order parameters we constructed display different timescales (10ps to 1ns), multiple timesteps can be implemented to further optimize the procedure and a larger average timestep for order parameters can be obtained to accelerate the simulations.

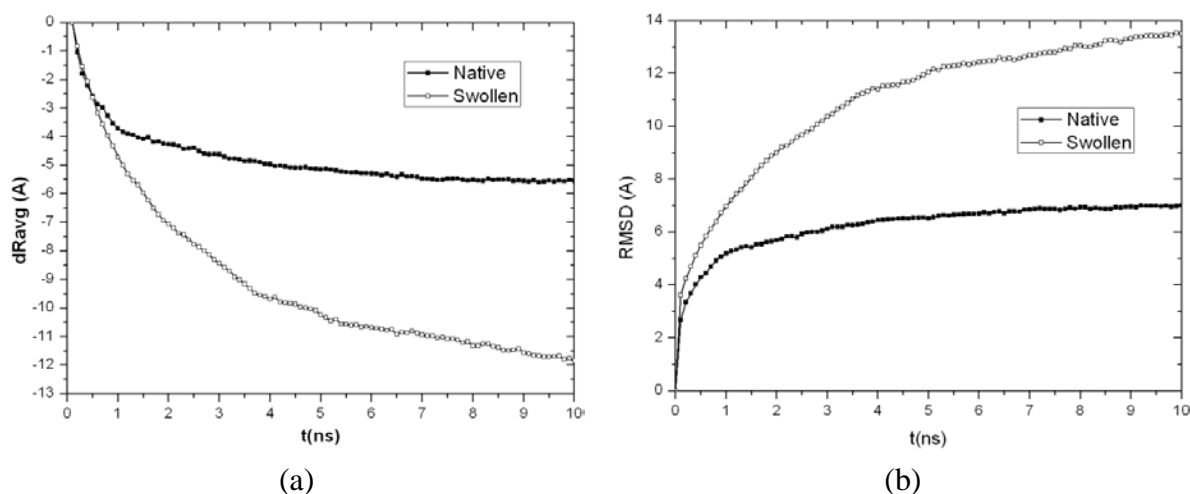
## **Chapter 8 Viral Capsid Structural Transition Mechanisms Revealed via Long-time MD/OPX Simulation**

### **8.1 Insights into CCMV Capsid Stability through Short-Time MD Simulations**

To identify key elements for carrying out long-time MD/OPX simulation of CCMV capsid, NAMD was firstly implemented to run short-time MD simulations (10ns) on its native and swollen states for insights into their structural stability. VMD[128] was used to create the capsid structural topology and NAMD[75] was implemented to simulate the native and swollen states of CCMV capsid in vacuum by using CHARMM22 force field [129, 130] for capsid proteins. Both simulations started with an initial energy minimization for 10,000 steps using the conjugate gradient algorithm, with which the system potential energies decrease to local minima and stay constants. Two energy-minimized structures were then gradually heated to 298.15K and equilibrated under this temperature for 20ps. Product MD runs capturing 10ns dynamics of the two capsid states were obtained to investigate their stability. A multiple-time-stepping algorithm[75] was employed for the MD simulations: bonded interactions were computed for every timestep, short-range nonbonded interactions every two timesteps and long-range electrostatic interactions every four timesteps.

With 10ns MD trajectories of the two structures obtained, frames taken for every 0.1ns were superimposed to their simulation starting configurations respectively to remove overall translation and/or rotation of the capsid. Then change in the capsid size was analyzed by calculating the decrement of capsid average radius, i.e., the average distance of non-hydrogen

atoms in capsid backbone to their center of mass (COM), from their initial values. The calculation results plotted in Figure III.7a show that the native state of CCMV capsid shrinks during the first 3ns of simulation with its average radius decreased by 4.6Å and becomes stable afterwards, while its swollen state shrinks much more significantly with the capsid average radius decreased by 11.8Å along the trajectory and still keeps shrinking at the end of the simulation at 10ns.



**Figure III.7** 10ns MD simulation of the native and swollen states of CCMV capsid: (a) time courses of the decrement in capsid average radius from their initial structures and (b) RMSD of atomic positions between their trajectory snapshots and the initial structures.

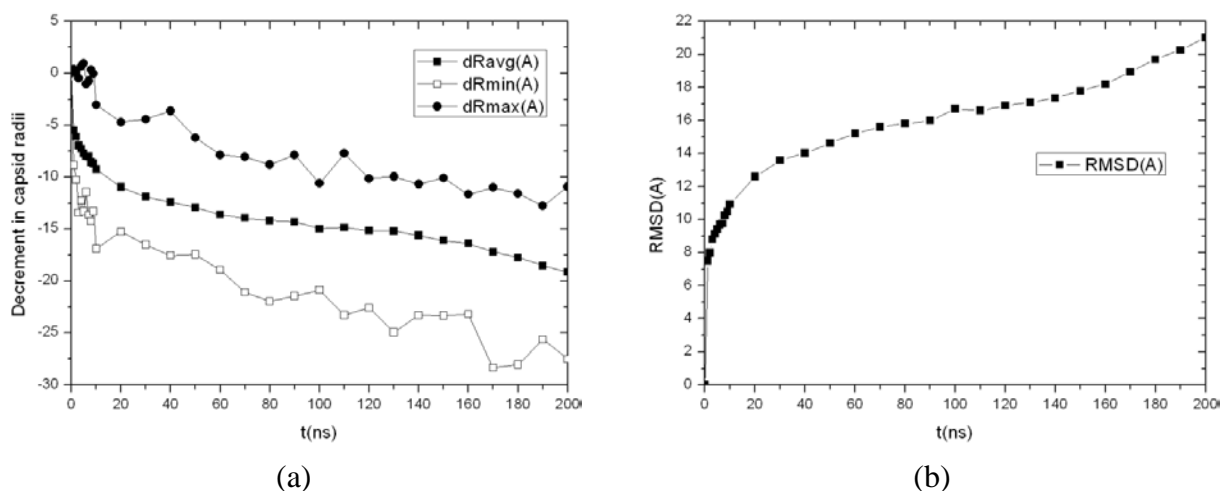
Structural changes in two states of the capsid can also be verified by observing time courses of the root mean square deviation (RMSD) of atomic positions between their trajectory snapshots and the initial configurations as shown in Figure III.7b. RMSD for the capsid native state is found to increase from 0Å to 6Å during 3ns and level off afterwards, while RMSD for its swollen state increases more dramatically (8.4Å at 3ns) and still displays increasing trend at 10ns. Therefore, native CCMV capsid tends to be stable after short, small shrinkage in vacuum,

while its swollen state undergoes shrinkage over a time period that is expected to be much longer than 10ns. With this, MD/OPX is implemented to simulate the long-time dynamics of swollen CCMV capsid in the following.

## **8.2 Shrinkage of Swollen CCMV Capsid in vacuum Captured by Long-time MD/OPX Simulation**

MD/OPX simulation of swollen CCMV capsid started with the resultant structure of its MD simulation at 1ns as described above and proceeded with cycles of short replica MD runs and projection of the atomic structure by extrapolating its ensemble-averaged order parameters over long time. In each cycle, ten 500fs replica MD runs with random atomic velocity initializations under 298.15K were obtained for computing the ensemble-average rates of change in capsid order parameters, which are then used to extrapolate the order parameters over long time. The newly obtained order parameters were taken to reconstruct atomic configuration of the capsid that will go through 1000-step energy minimization, 6-stepwise graduate heating to 298.15K, and 5ps MD equilibration. Atomic accelerations of the resulting structure were computed as indicators to determine whether the structure is mature for the next MD/OPX cycle and the timestep for extrapolation of order parameters is adjusted adaptively to ensure the simulation is stable. By repeating such cycles, swollen CCMV capsid was simulated for 200ns in vacuum with its shrinkage captured. The OP timestep obtained from adaptive MD/OPX cycles was found to undergo periodic oscillations and its average for the entire simulation was determined to be 60ps. With this, the ratio of average OP timestep to the time interval of short MD bursts is

largely increased to 120, compared with 10 obtained from the MD/OPX demonstration simulation by using only a single short MD run for OPX (see Chapter 7) [12]. When comparing the simulation time of MD/OPX to that of one single MD run, the overall speedup of MD/OPX over MD is by a factor of 6. However, as ten short replica MD runs were used in MD/OPX for extrapolating the order parameters, ensemble average effects have been accounted for and our MD/OPX becomes 60 times faster than an ensemble of 10 replica MD runs. The simulation was run on Indiana University Big Red cluster with performance of 0.06 days/ns when using 256 processors and 0.11 days/ns for 128 processors.



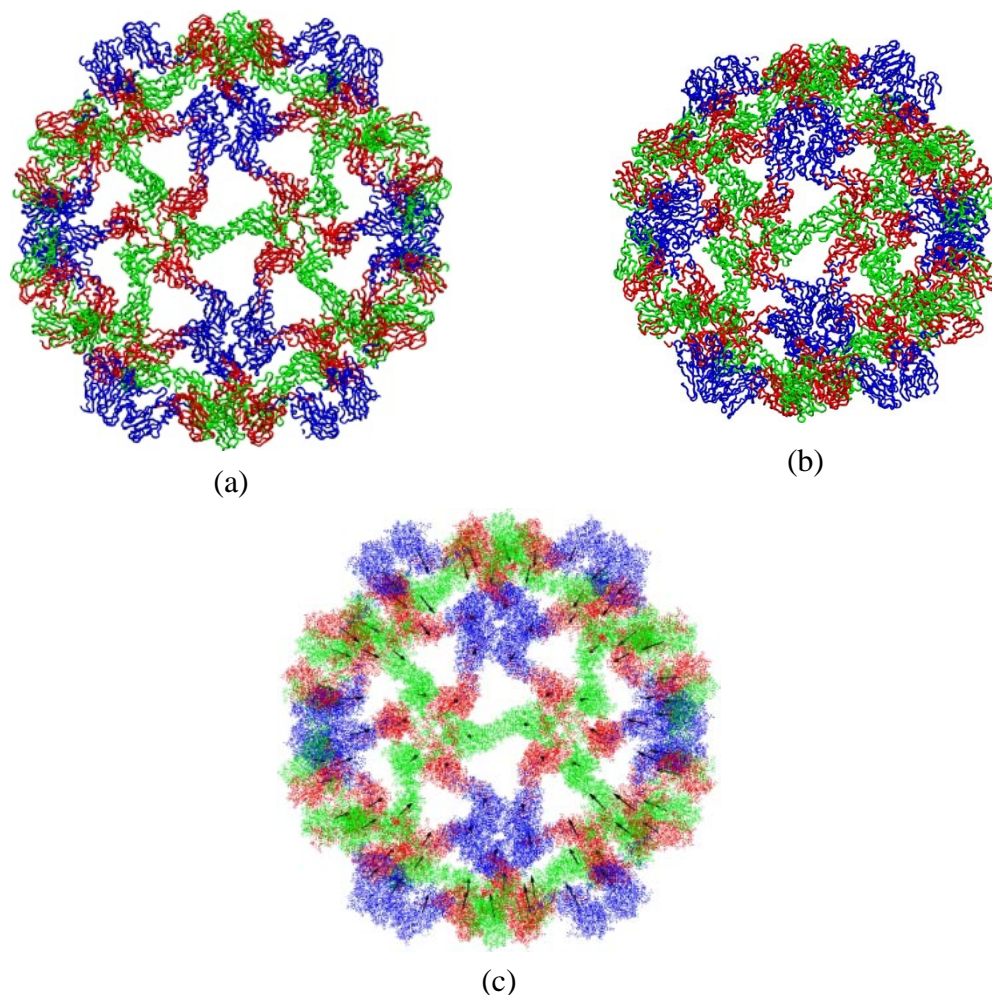
**Figure III.8** 200ns MD/OPX simulation of the swollen CCMV capsid: (a) time courses of the decrement in the average, minimum and maximum radii of the capsid backbone from the simulation starting structure, and (b) RMSD of atomic positions between capsid snapshots along the trajectory and the initial structure.

Analysis on the simulation output trajectory shows that swollen CCMV capsid shrinks with the average radius of its backbone decreased by  $19.15\text{\AA}$  during 200ns (Figure III.8a). It thickens by  $11\text{\AA}$  during shrinkage and undergoes fluctuations as indicated by the time courses of the



decrement in the minimum and maximum radii of the capsid backbone (i.e., the minimum and maximum distances of capsid backbone atoms to the capsid COM). RMSD of atomic positions between capsid snapshots along the trajectory and its starting structure increases from  $4.10$  to  $21.01\text{\AA}$  in 200ns (Figure III.8b). The capsid is found to shrink fast in early stage of the simulation with quick drop in the capsid average radius and sharp increase in the capsid RMSD, and the shrinkage slows down as the capsid approaches its near-equilibrium state that is expected to be close to the native.

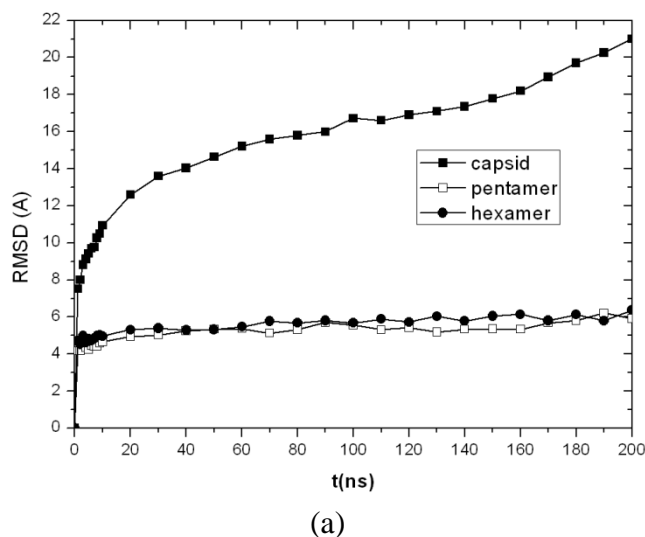
To identify characteristics of the shrinkage of swollen CCMV capsid in vacuum, interior view of its back half (selected through  $z < 0$  with the capsid centered at origin) for the simulation starting and final structures are shown in Figure III.9 (a) and (b). In Figure III.9c, arrows are drawn from subunit COMs of the capsid initial structure to their final positions. It can be observed that pentamers and hexamers in the capsid translate radially inwards and rotate in a clockwise manner during capsid shrinkage. As a result, the openings between the capsomeres along the quasi-threefold axes become small heading towards the “closed” native state. This acts as a reverse process to the pH-induced expansion of native CCMV capsid into its swollen state in host medium.

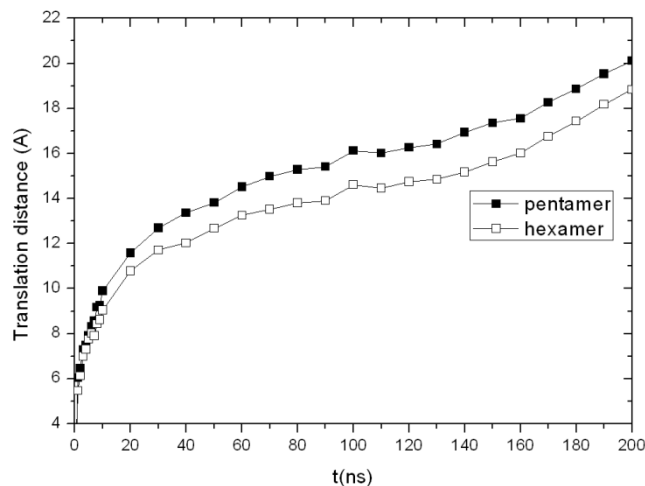


**Figure III.9** Interior views of (a) swollen CCMV capsid, (b) the final structure of 200ns MD/OPX simulation, and (c) the COM displacements of protein subunits from their initial configurations to the final that depict shrinkage of the capsid.

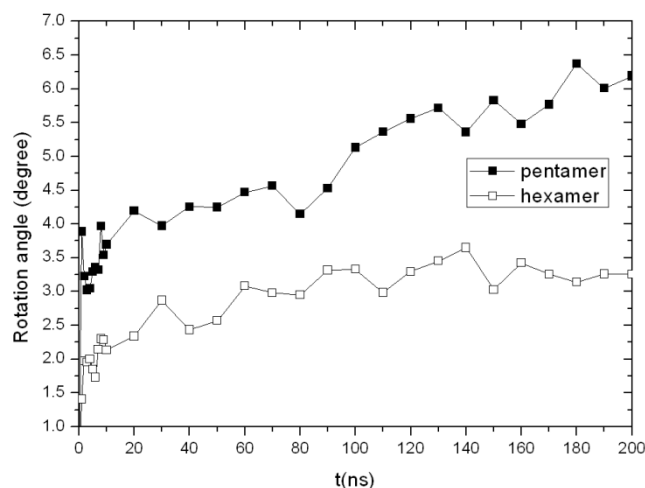
To examine the suitability of considering pentamers and hexamers as rigid units during CCMV capsid shrinkage, we choose a pentamer and a hexamer from the capsid, calculate their RMSD of atomic positions between trajectory snapshots and their initial configurations after structure superimposition, and compare them to that for the entire capsid. Results in Figure III.10a show that RMSD for a pentamer or a hexamer increases to  $\sim 1$  Å in 1 ns and displays slight changes afterwards, while RMSD for the entire capsid keeps increasing through the simulation as described above. This indicates that the pentamer and hexamer adjust their

internal structures quickly during early stage of the simulation and become stable afterwards during shrinkage of the capsid, which then can be considered as a process of rearranging the capsid pentamers and hexamers by overall translation and rotation. Calculations on the average translation distances and rotation angles of 12 pentamers and 20 hexamers during capsid shrinkage give us quantitative results as follows. As shown in Figure III.10 (b) and (c), pentamers translate radially inwards by  $20.11\text{\AA}$  on average during 200ns and hexamers translate by  $18.84\text{\AA}$ , and they undergo clockwise rotation by about  $6.19^\circ$  and  $3.26^\circ$  respectively with moderate fluctuations. The simulated capsid shrinkage transforms pentamers and hexamers in directions that are reverse to those applied to generate the swollen capsid from its native state.





(b)

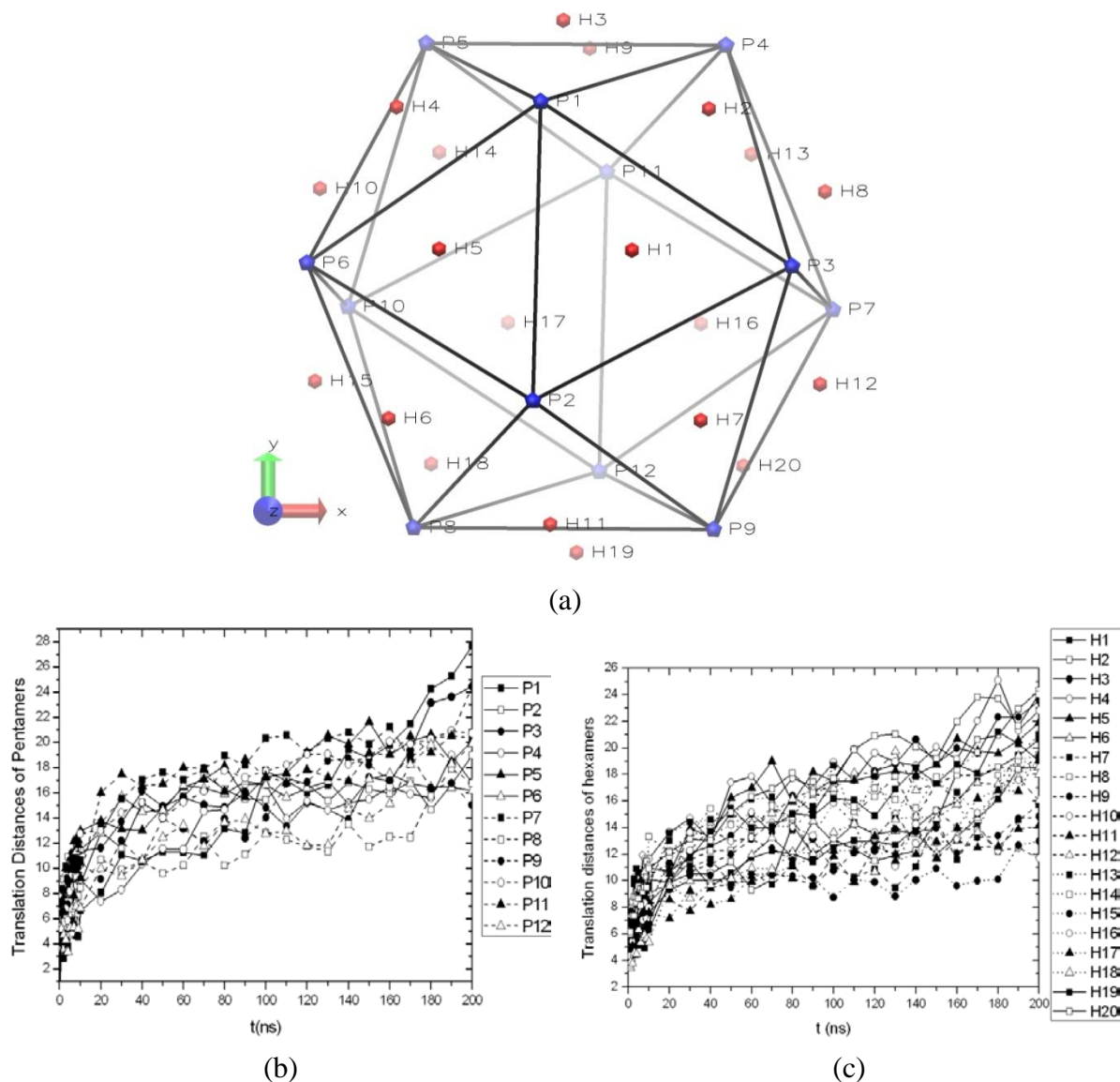


(c)

**Figure III.10** (a) RMSD of atomic positions for a selected pentamer and a hexamer (see Figure III.11a for P1 and H1) between trajectory snapshots and their initial configurations in comparison to that for the entire capsid, (b) the average COM translation distance as a function of time for pentamers and hexamers, and (c) time courses of the average rotation angle for pentamers and hexamers calculated through fitting their structures to the initial configurations.

While above results are obtained by averaging translation distances and rotation angles over the pentamers and hexamers, a further step is to look into the transformation of capsomeres individually during capsid shrinkage. To do this, 12 pentamers from P1 to P12 and 20 hexamers from H1 to H20 are first labeled in icosahedral CCMV capsid according to the scheme provided

in Figure III.11a. Then their COM translation distances from the initial positions are plotted as a function of time in Figure III.11 (b) and (c) for pentamers and hexamers respectively. Results show that the 12 pentamers do not translate at the same speed with P11 and P7 moving the fastest, P8 and P12 the slowest, and the others between through most of the simulation time. Notice that P1 and P3 lead the translation of pentamers from 170ns to the end of simulation. For hexamers, H13, H20, and H4 are in the fast moving group, while H15 and H17 belong to the slowly-moving. By examining the distribution of these capsomeres in Figure III.11a, one can find that the fast moving hexamers and pentamers tend to be neighboring capsomeres, e.g., H13 to P11 and H20 to P7, and similarly for the slowly-moving ones, e.g., H15 to P8 and H17 to P12. This implies that the motions of capsomeres are largely correlated and it results from the structural features of CCMV capsid. According to CCMV structural studies [131], the carboxyl-terminal arms of protein subunits in its pentamers and hexamers are known to “invade” their neighboring capsomeres, through which the capsomeres are tightly intertwined with each other. Thus capsomeres would undergo cooperative motions with strongly-coupled allosteric interactions. As a result, swollen CCMV capsid does not preserve its icosahedral symmetry during shrinkage and it involves symmetry-breaking cooperative motions of the capsomeres.



**Figure III.11** (a) Schematic representation of icosahedral CCMV capsid with 12 pentamers (blue pentagons) labeled from P1 to P12 and 20 hexamers (red hexagons) from H1 to H20, and time courses of the COM translation distance of (b) pentamers and (c) hexamers from their original positions.

### 8.3 Mechanisms of Viral Capsid Structural Transition

As indicated above, shrinkage of swollen CCMV capsid in vacuum involves large-scale rearrangements of the pentamers and hexamers and icosahedral symmetry is not preserved during the transition. Symmetry-breaking is a well-known feature in the self-organization and

structural phase transitions of macroscopic systems [132, 133]. For example, martensitic transformations that are characterized by a collective movement of large numbers of atoms occur in solid-state materials, such as iron and metal alloys, and they are found to change system symmetries [134, 135].

It has been suggested that STs in viral capsids are analogous to structural phase transitions in macroscopic solids and thus methods in solid-state physics can be applied to viral capsids. Based on this hypothesis, viral capsid STs were studied with phenomenological continuum theory adapted from the Ginzburg-Landau model of soft-mode solid structural phase transitions. The theory predicts that the STs are characterized by a pronounced softening of the capsid elasticity and external force applied by an AFM probe can drive a capsid into a state of phase coexistence where the capsid is partly in the immature state and partly in the mature state with two of them separated by a phase boundary[136].

To understand the mechanisms of viral capsid ST, allosteric coupling has been proposed by Caspar [137]. It states that the conformational change in one capsid subunit triggers that in neighboring subunits, which then induces structural changes in more subunits across the capsid. Such cooperative conformational changes display symmetry-breaking behaviors as found for pentamers and hexamers in swollen CCMV capsid during its shrinkage. This is common to STs of many other viral capsids during maturation as well [138, 139].

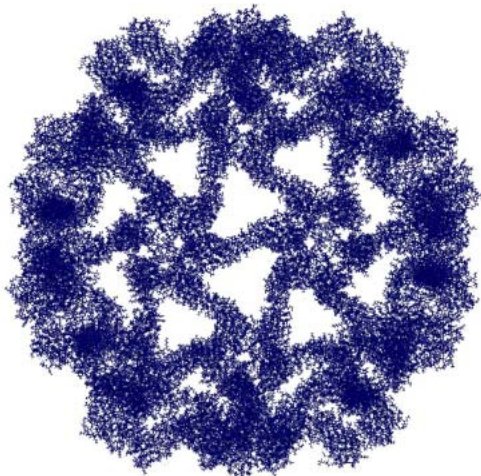
Experimental and theoretical studies on HK97 bacteriophage capsid show that it undergoes expansion from a metastable procapsid (prohead) to a mature icosahedral capsid (head) via three steps[51]: 1) local refolding of capsid subunit N-terminal arms, 2) global expansion of the

capsid involving large-scale rearrangements of capsid subunits, and 3) additional subtle and slow structural changes contributing to cross-linking of capsid subunits that stabilize the mature capsid. While greatest structural changes take place during capsid global expansion, they are triggered by local refolding of the capsid subunit N-arms. This suggests that the capsid does not preserve its symmetry during the initial highly-localized conformational changes and even the cooperative motions of capsid subunits during its expansion [139]. MD simulations on satellite tobacco mosaic virus (STMV) [52] show that its capsid also undergoes symmetry-breaking ST. While the complete virion (capsid embedded with RNA) deviates its shape from icosahedral symmetry with only minor, local ones over 13ns of simulation, empty viral capsid was found to collapse and its shape becomes very different from the initial state after 10ns with the icosahedral symmetry lost completely.

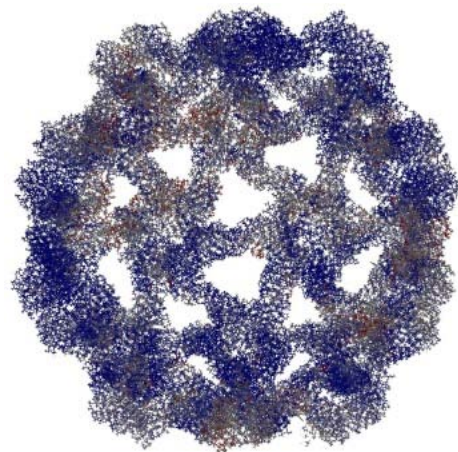
While pH-induced CCMV swelling has been determined to be a first-order transition as suggested by hysteresis of CCMV in its titration curves under certain conditions [126] and pH-driven softening of capsid elasticity in AFM experiments[140], exact mechanisms of the capsid ST remain unclear [45]. With our all-atom multiscale MD/OPX simulation of swollen CCMV capsid, we are able to obtain a microscopic view of its STs, and therefore explore the exact symmetry-breaking mechanisms during capsid shrinkage. Figure III.12 shows interior view of the capsid trajectory snapshots at (a) initial swollen state, (b) 5ns, (c) 20ns, (d) 50ns, (e) 100ns, and (f) 200ns. Atoms in the capsid are colored by their displacements from the original positions in a blue-white-red color scale (i.e., 0 Å for blue, 24 Å for red, and white as the midpoint) to investigate conformational changes in the swollen CCMV capsid. It is seen that



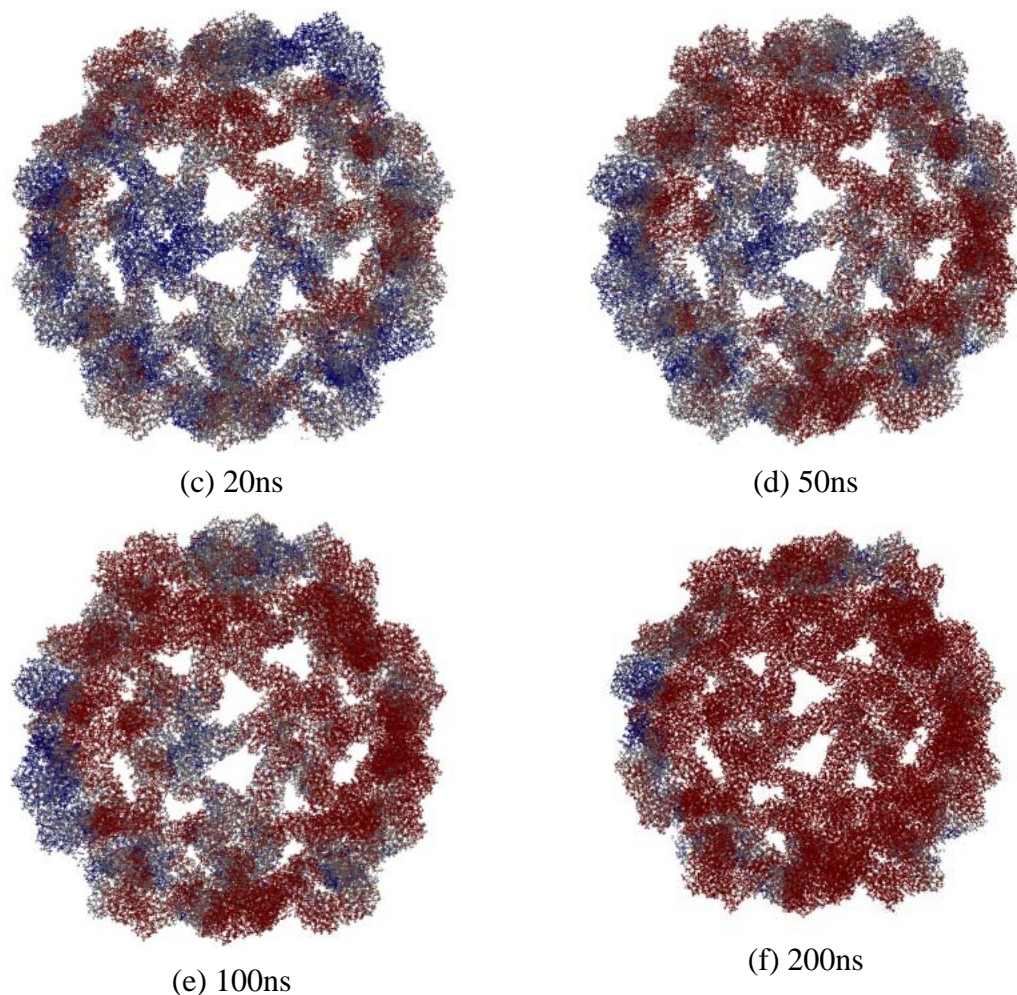
atomic displacements are not constrained to icosahedral symmetry, i.e., atoms in the capsid do not change their colors simultaneously in a symmetry-preserving manner. Shrinkage of swollen CCMV capsid involves an “initiation and front propagation” mechanism: it starts from the upper pentamer as shown in trajectory snapshots and several other local regions of the capsid in which atoms change their colors from blue to white and to red first. This nucleation needs 20ns to occur. The structural transition then propagates across the capsid seen as spreading of the red areas. The propagation of cooperative motions from one capsomere to its neighboring ones takes place on a timescale of 10ns. Thus the propagation speed is  $\sim 0.6\text{nm/ns}$  with the average capsomere distance considered as 6nm. Therefore, STs of viral capsids start with local conformational changes that break icosahedral symmetry of the capsid, followed by global structural changes that involve large-scale cooperative translations and rotations of capsid structural units (e.g., capsomeres or protein subunits) through strong allosteric interactions.



(a) initial



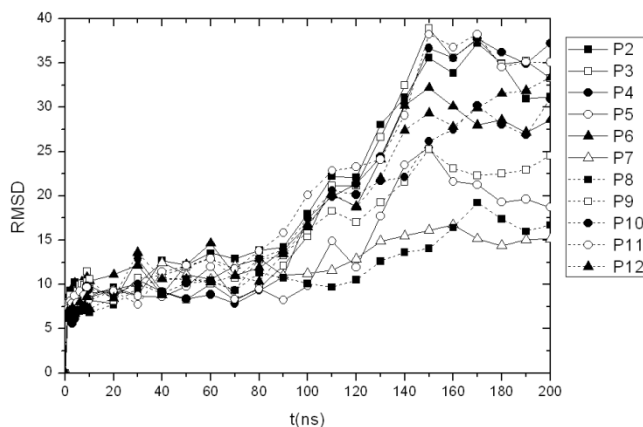
(b) 5ns



**Figure III.12** Interior view of the capsid shrinkage trajectory snapshots with atoms colored by their displacements from the original positions in a BWR color scale ( $0\text{\AA}$  for blue,  $24\text{\AA}$  for red, and white as the midpoint): (a) initial state with all atoms in blue, (b) 5ns, (c) 20ns, (d) 50ns, (e) 100ns, and (f) 200ns of simulation. It can be seen that atoms in the upper pentamer of shown trajectory snapshots (i.e., P11 as labeled in Figure III.11a) start to change their color from blue to white and to red first, and then this color change propagates across the capsid. Thus capsid shrinkage is a symmetry-breaking/front propagation process.

To evaluate symmetry-breaking of CCMV capsid during its shrinkage quantitatively, we calculate the RMSD of atomic positions between pentamer 1 as labeled in Figure III.11a and the 11 others after rigid-body transformation of the chosen pentamer to the positions of others according to icosahedral symmetry. These RMSDs (Figure III.13) are found to increase up to

about 140ns with structural changes breaking capsid symmetry, and then decrease afterwards as the capsid transforms towards its near-equilibrium state (one close to the icosahedral native state). As a result, even though the initial and final states of swollen CCMV capsid have icosahedral symmetry, symmetry is not preserved during its shrinkage.



**Figure III.13** Time courses of RMSD of atomic positions between pentamer P1 as labeled in Figure III.11a and the 11 others after rigid-body transformation of P1 to the positions of other pentamers according to icosahedral symmetry.

The fact that shrinkage of swollen CCMV capsid in vacuum is symmetry-breaking at first seems to conflict with prevailing thinking. For example, symmetry has been used to facilitate the simulation of icosahedral viruses by assuming symmetry is preserved during their STs [53, 69, 72]. However, as with structural phase transitions in macroscopic solids, viral capsid STs start locally and then propagate across the capsid, i.e., they proceed via intermediate states that are not constrained to the capsid symmetry[134]. Examples of such viral symmetry-breaking STs include those induced by local interactions of viral capsids with cell receptors or antiviral drug molecules [7, 43, 60, 127]. Given the fact that viral capsid STs are symmetry-breaking

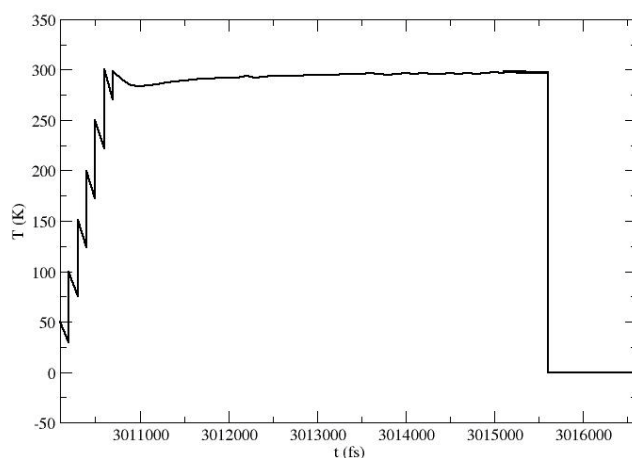
processes, the use of symmetry-constrained models in studying the pathways and mechanisms of STs in CCMV and other icosahedral viruses is not appropriate. Particularly, the energy barrier for making coherent translations and rotations of all capsomeres or capsid subunits in the course of viral capsid STs is much greater than that for a single or local cluster of these substructures. Thus one expects that prediction of transition time based on a symmetry-constrained model should be unphysically long.

#### **8.4 Shrinkage of Swollen CCMV Capsid is Energy-driven**

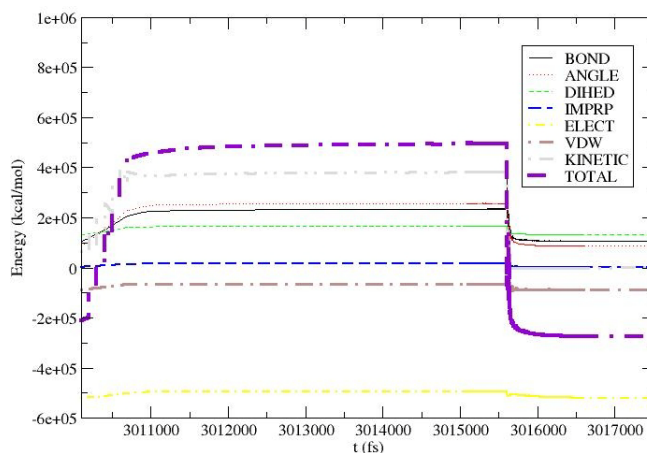
In the following, we examine system energies for shrinkage of swollen CCMV capsid in vacuum simulated via MD/OPX. Since MD/OPX cycles are composed of short replica MD runs, projection of system configuration through extrapolation of its order parameters and maturation of the OPX resulting structure by applying energy minimization, graduate heating and short-time MD equilibration (see Experimental Procedures), and temperature control is applied all through the simulation, we first look into variations of the system temperature and energies in one typical MD/OPX cycle.

For our analysis, the 32nd MD/OPX cycle of the simulation running from 3,010,100fs to 3,075,100fs with the ratio of OP timestep to the time interval of short MD runs obtained as 120 (also the average value for the entire simulation) was chosen. In Figure III.14a, the system temperature was plotted as a function of time: it starts from 0K as a result of energy minimization of the OPX resulting configuration from previous cycle; increases to ~298.15K in a 6-stepwise manner as the system is heated gradually by 50K for each of the first 5 steps and

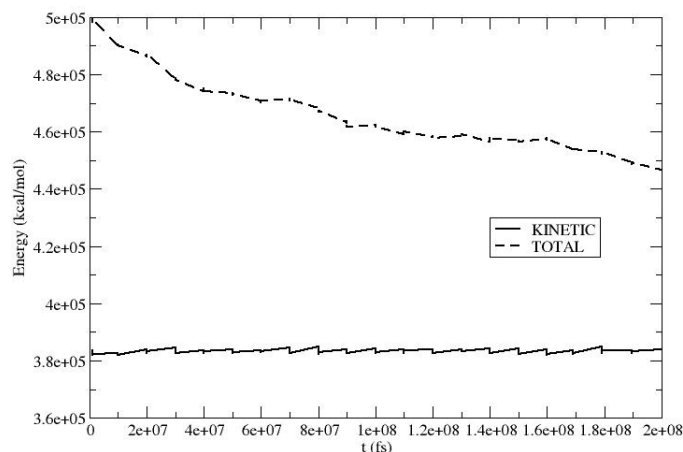
48.15K for the last step (100fs equilibration for each temperature increase); becomes stable as the system is equilibrated to 298.15K for 4400fs; stays around 298.15K under which ten 500fs short replica MD bursts are run for calculating the ensemble-averaged rate of change in system order parameters; and drops to 0K as the atomic configuration is reconstructed with newly extrapolated order parameters over 60, 000fs and energy minimization is performed. Note that for each heating step, atomic velocities are reinitialized for the system to reach a higher temperature, after which there is a slight drop in system temperature due to conversion of the system kinetic energy into potential energies as seen below.



(a)



(b)



(c)

**Figure III.14** Variations of the system temperature (a) and energies (b) in one typical MD/OPX cycle capturing 3,010,100-3,075,100fs of the simulation, and (c) the system kinetic and total energies plotted versus time during 200ns capsid shrinkage.

In Figure III.14b, system energies for the 32nd MD/OPX cycle were plotted versus time corresponding to the system temperature shown in Figure III.14a. Starting from comparatively low values computed for the energy-minimized configuration of previous cycle, bonded energies (i.e., bond, angle, dihedral and improper torsion energies) were found to increase significantly during stepwise heating of the system, while there are slight increases to the system non-bonded energies (i.e., electrostatic and van der Waals energies). As a result, heat supply from thermal bath (i.e., increase in system kinetic energy) is converted into system potential energies and system total energy is increased. After the graduate heating procedure, capsid energies level off as the system is equilibrated and stay nearly constants during short replica MD runs that are used for OPX. As for the OPX resulting configuration, it can lead to high potential energies, especially for non-bonded van der waals and electrostatic energies, which are due to unphysical structures, such as overlapped atoms and stretched bonds/angles. They are seen as spikes in energy curves during early stage of system energy minimization.

These spikes disappear and system potential energies decrease significantly as more energy minimization steps are applied, which prepare the system atomic configuration to be ready for MD thermalization of next cycle.

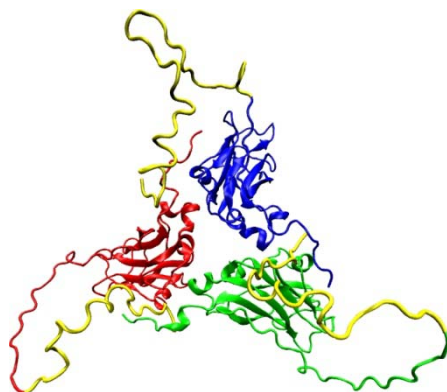
For the entire 200ns MD/OPX simulation, kinetic and total energies of swollen CCMV capsid were collected approximately every 10ns from the short replica MD runs of MD/OPX cycles and plotted as shown in Figure III.14c. The system kinetic energy was found to stay constant with small fluctuations, which suggests that the graduate heating and MD equilibration are sufficient to thermalize the energy-minimized OPX resulting configuration for running short replica MD runs and thus computing ensemble-average rate of change in system order parameters for extrapolation. On the other hand, the system total energy was shown to decrease monotonically from 498,406kcal/mol to 447,539kcal/mol. This indicates that the system potential energy decreases accordingly along the trajectory. As the shrinkage of swollen CCMV capsid mostly involves large-scale rearrangements of its pentamers and hexamers through cooperative motions as described above, the system entropy is expected to stay roughly constant, and thus CCMV capsid shrinkage is an energy-driven process.



## Chapter 9 All-atom Multiscale MD/OPX Simulation of CCMV Capsid Swelling

### 9.1 Insights into CCMV Capsid Swelling through Short-Time MD Simulations

Similarly, to identify key elements needed to carry out long-time MD/OPX simulation of CCMV capsid swelling, short-time MD simulation was first run on native CCMV capsid that is solvated in a host medium at pH 7.0 and ionic strength  $I=0.2\text{M}$ . The X-ray crystal structure of CCMV protomer (i.e., asymmetric unit) was downloaded from online Protein Data Bank (PDB ID: 1CWP) and the missing N-terminal residues of protein subunits were added and optimized (refer to the scheme provided in Zhang et al. [73]) (Figure III.15).

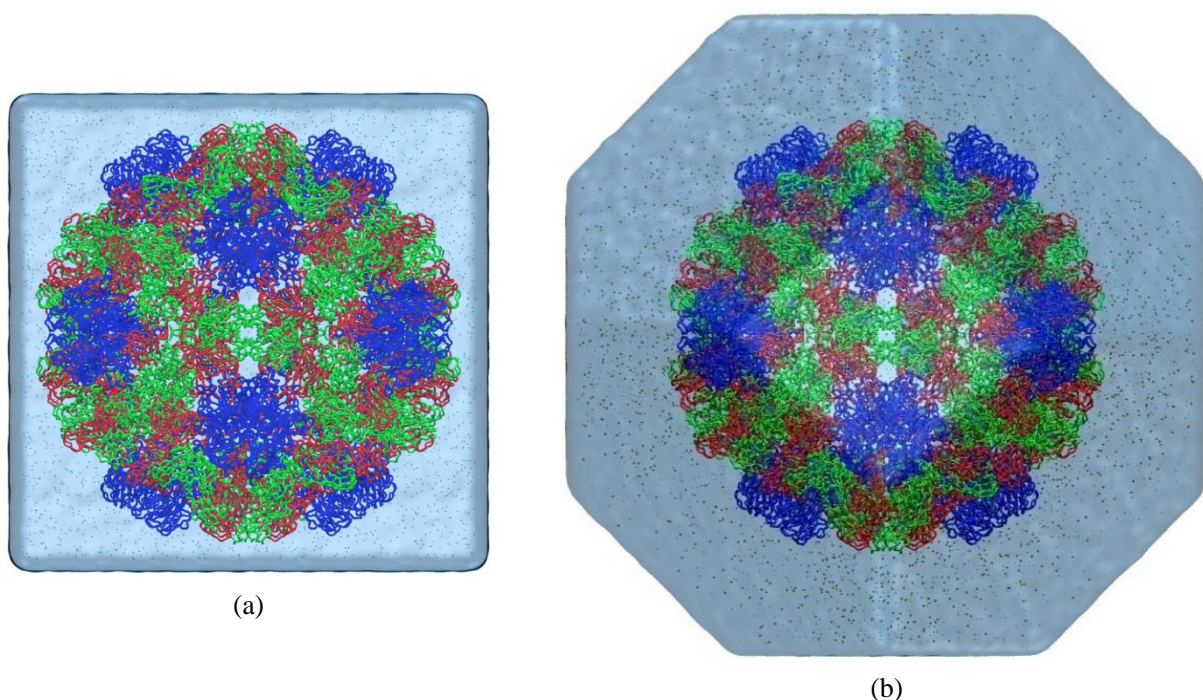


**Figure III.15** X-ray crystal structure of CCMV protomer with the missing N-terminal residues added and optimized. Three quasi-threefold related protein subunits are colored in blue (A), red (B) and green (C) respectively, and the N-terminal arms in yellow.

Native CCMV capsid was generated by reorienting 60 copies of its protomer (i.e., asymmetric unit) according to the icosahedral symmetry and solvated in the center of a water cube with  $7\text{\AA}$  boundary in each direction and neutralized with  $\text{K}^+$  and  $\text{Cl}^-$  to  $I=0.2\text{M}$  by using VMD. A schematic representation of the resulting system is shown in Figure III.16a. Then the system along with its structural topology file were taken as input for NAMD[75] simulation by



using the CHARMM22 force field[129, 130] for capsid proteins and TIP3P model for water molecules[141]. The simulation proceeded with initial energy-minimization, thermalization with graduate heating and MD equilibration under constant temperature (298.15K) and pressure (1atm) control (i.e., NPT ensemble). The system was equilibrated after 20ps when the system temperature and pressure fluctuate around target values. Product MD run lasting 3ns was obtained to investigate structural changes of the capsid to identify essential features needed to model swelling of the capsid.



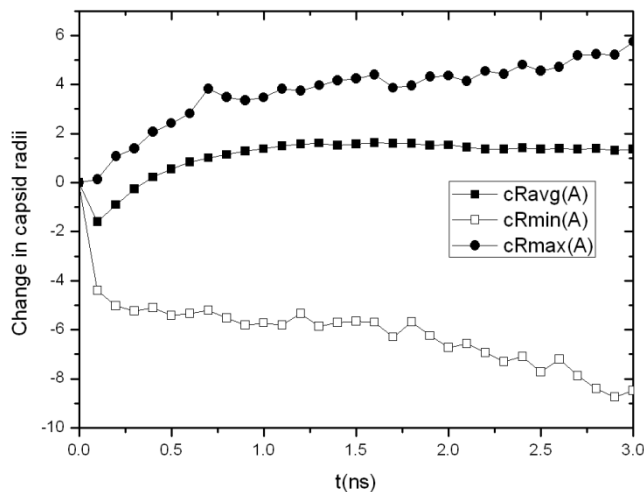
**Figure III.16** Schematic representations of CCMV capsid solvated in (a) a water cube and (b) a water truncated octahedron at pH 7.0 and ionic strength  $I=0.2\text{M}$  used for our simulations with the water box created using VMD “volmap” plugin and ions represented in CPK.

Analysis of the MD trajectory shows that the capsid shrinks by  $1.60\text{\AA}$  in the average radius of its backbone (i.e., the average distance of atoms in the capsid backbone to their COM) as a result of initial energy minimization, thermalization, and MD equilibration. Especially during

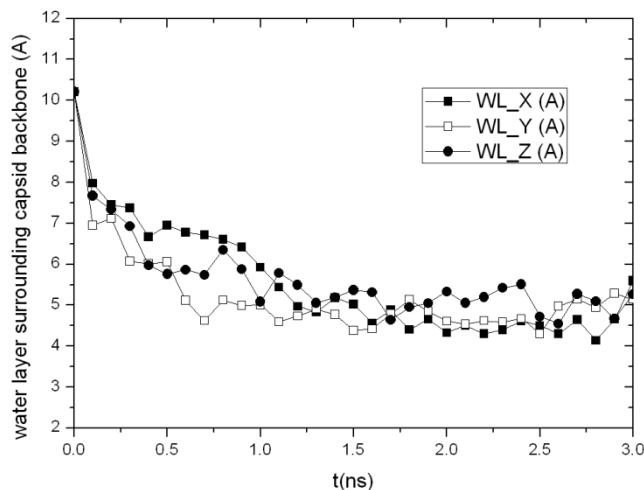
the MD equilibration under pressure control, the capsid is compressed with the entire system scaled isotropically to reach 1atm pressure. The capsid then starts to swell after equilibration. For the resulting structure at 3ns, its average radius is 2.95Å greater than the equilibrated structure and 1.35Å greater than the native CCMV capsid. It was also found to be 13.99Å thicker than the native state with its maximum radius (i.e., the maximum distance of capsid backbone atoms to the capsid COM) increased from 142.49Å to 148.23Å and the minimum radius decreased from 82.32Å to 73.85Å (Figure III.17a). The thickness of water layer surrounding the capsid backbone in X, Y and Z directions, which are calculated by averaging the minimum distance between non-hydrogen atoms in the capsid backbone (excluding those in protein side chains) and the faces of water box on two sides of each direction, were found to decrease as the capsid swells (Figure III.17b). As a result, the distance between the capsid surface and that of its replicas in neighboring periodic cells decreases. When it reaches the cutoff distance of non-bonded atomic interactions (usually 12Å for short-range van der Waals and electrostatic interactions in MD simulation), the capsid starts to interact with the replicas of its neighboring cells, which slows down swelling of the capsid.

The above MD simulation shows that native CCMV capsid swells in 3ns when placed in a host medium at pH 7.0 and ionic strength  $I=0.2M$ . However, the simulation takes 2.60 days to finish 1ns by using 256 processors on Indiana University Big Red cluster. And as the capsid swells, the water box needs to be expanded to appropriately simulate swelling of the capsid (see discussions below). With the system size largely increased, the performance of MD simulation is downgraded and it becomes even more time-consuming to simulate the swelling of CCMV

capsid. Since viral STs involve processes taking place across multiple time and length scales, we will apply multiscale MD/OPX to CCMV capsid swelling, which accelerates MD for long-time simulation and addresses the atomistic fluctuations and nanosystem coherent motions simultaneously. Simulation results are presented in following.



(a)



(b)

**Figure III.17** 3ns MD simulation of native CCMV capsid solvated in a water cube with an initial 7Å water boundary: (a) changes in the average, minimum and maximum radii of the capsid backbone from the starting configuration, (b) averaged thickness of water layer surrounding the capsid backbone in X, Y and Z directions, and (c) RMSD of atomic positions between snapshots of capsid backbone along the trajectory and the starting configuration.

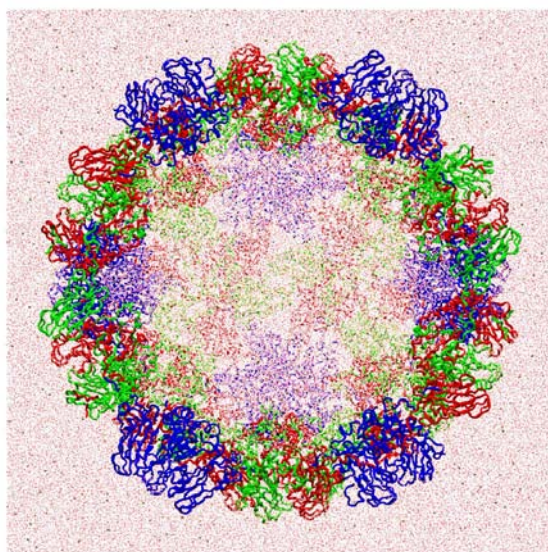
## 9.2 MD/OPX Simulation of CCMV Capsid Swelling

Since earlier MD/OPX studies [12, 13], new modules have been added to account for water molecules and ions and thus achieve simulating bionanosystems in host media. Here MD/OPX is used to simulate swelling of CCMV capsid at pH 7.0 and ionic strength  $I=0.2M$ . As native CCMV capsid is expected to expand by  $\sim 24\text{\AA}$  to reach the size of its swollen state[18], simulation of its swelling process is divided into the following two parts for efficiency: (1) increase the water boundary from  $7\text{\AA}$  used in the above MD simulation to  $15\text{\AA}$  for solvating native CCMV capsid (denoted Ncap) and run MD/OPX on the resulting system until the capsid swells to the size limit of the water cube, and (2) extract the intermediate capsid structure obtained from (1) (denoted Icap) and solvate it in a larger water cube for further simulation of the capsid swelling.

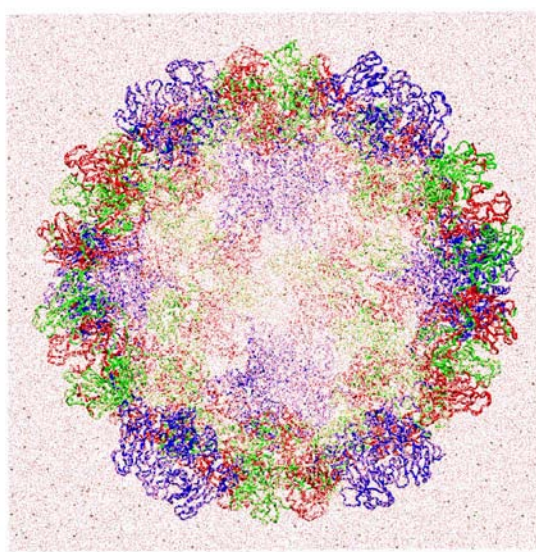
For the first part of the simulation, MD/OPX started with native CCMV capsid (Ncap) solvated in a water box with  $15\text{\AA}$  water boundary. The simulation proceeded with cycles of short MD runs and projection of the atomic structure by extrapolating its structural order parameters over long time. In each cycle, one 100fs replica MD runs with random atomic velocity initializations under 298.15K was obtained for computing the rates of change in capsid order parameters, which are then used to extrapolate the order parameters over long time. The newly obtained order parameters were taken to reconstruct atomic configuration of the capsid that will be resolvated in a host medium as described above and go through 5000-step energy minimization, 6-stepwise graduate heating to 298.15K, and 5ps MD equilibration. Atomic accelerations of the resulting structure were computed as indicators to determine whether the



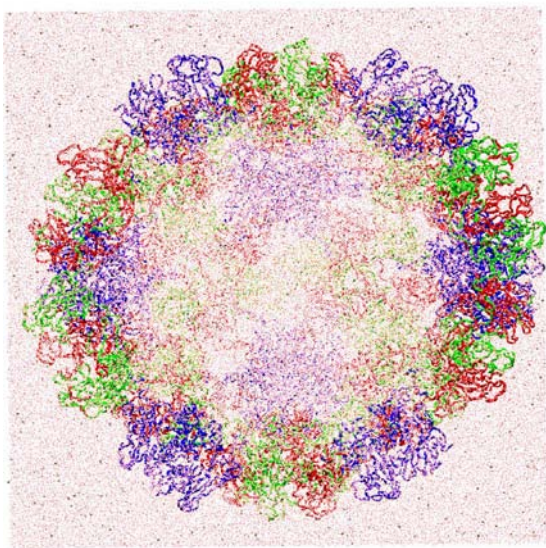
structure is mature for the next MD/OPX cycle and the timestep for extrapolation of order parameters is adjusted adaptively to ensure the simulation is stable. By repeating such cycles, the simulation capturing 3ns dynamics of the capsid was obtained to study its structural changes.



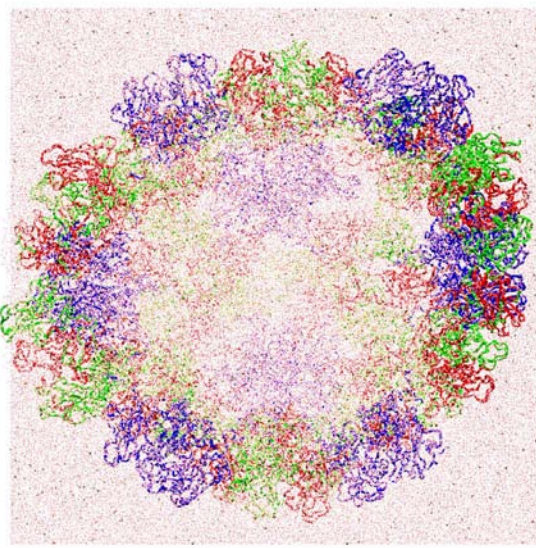
(a) Initial



(b) 0.5ns



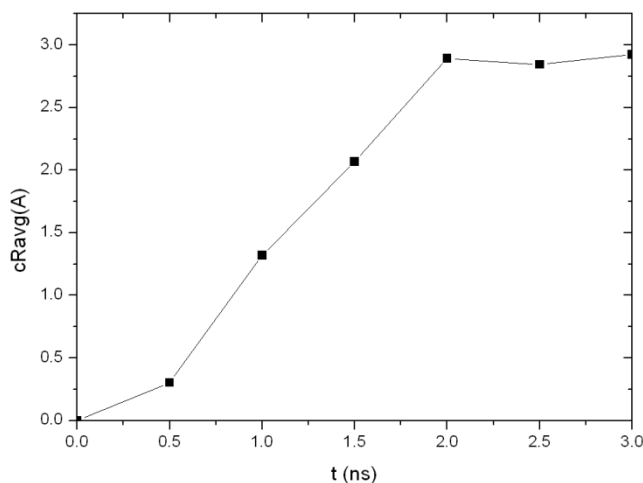
(c) 1.5ns

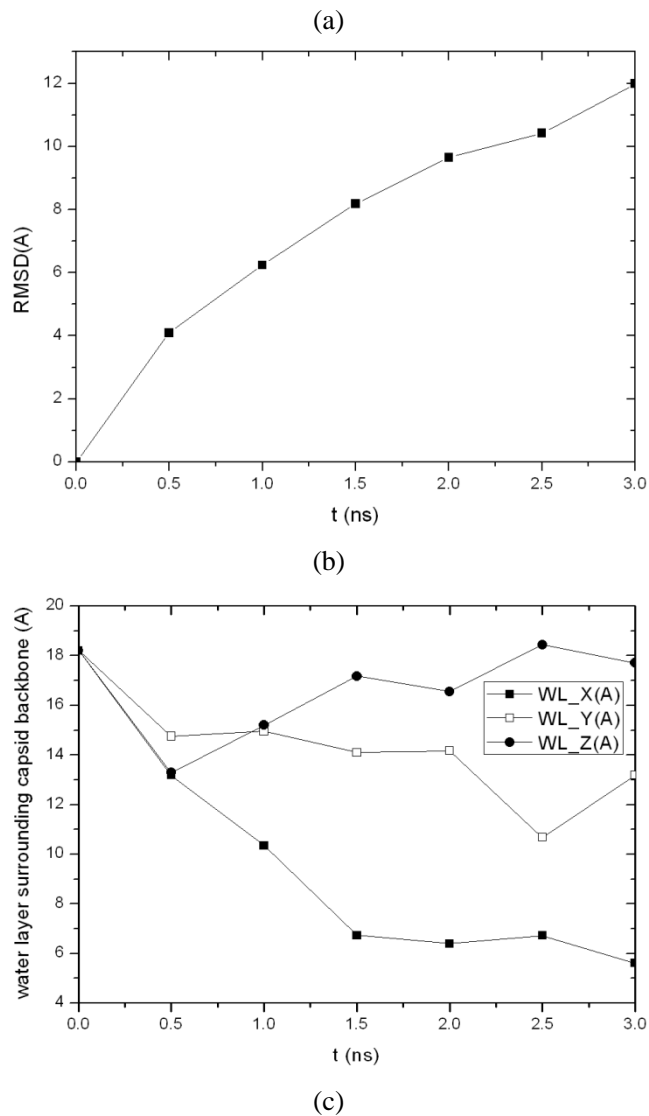


(d) 3ns

**Figure III.18** MD/OPX trajectory snapshots showing the expansion of native CCMV capsid (Ncap) in water cube with an initial 15Å water boundary: (a) the starting configuration, (b) 0.5ns, (c) 1.5ns and (d) 3ns.

Figure III.18 shows trajectory snapshots of the capsid in water cube at: (a) the starting configuration, (b) 0.5ns, (c) 1.5ns, and (d) 3ns. The capsid was seen to expand in the MD/OPX trajectory, which agrees with the above MD simulation. Quantitative analysis of the MD/OPX trajectory shows that the capsid swells with the average radius of its backbone increased by  $\sim 3\text{\AA}$  in 3ns (Figure III.19a). RMSD of atomic positions between capsid snapshots along the trajectory and its starting structure increases from  $0\text{\AA}$  to  $12\text{\AA}$  in 3ns (Figure III.19b). The capsid is found to swell fast up to 2ns, and the swelling slows down thereafter. Calculations show that the thickness of water layer surrounding the capsid backbone decreases as a result of the capsid swelling. It reaches  $6\text{\AA}$  (half of the cutoff distance for atomic non-bonded interactions) at 1.5ns, within which the capsid starts to interact with the replicas of its neighboring cells. Therefore, it is reasonable to extract the capsid structure from the trajectory snapshot at 1.5ns and solvate it in a larger water cube for further swelling simulation.

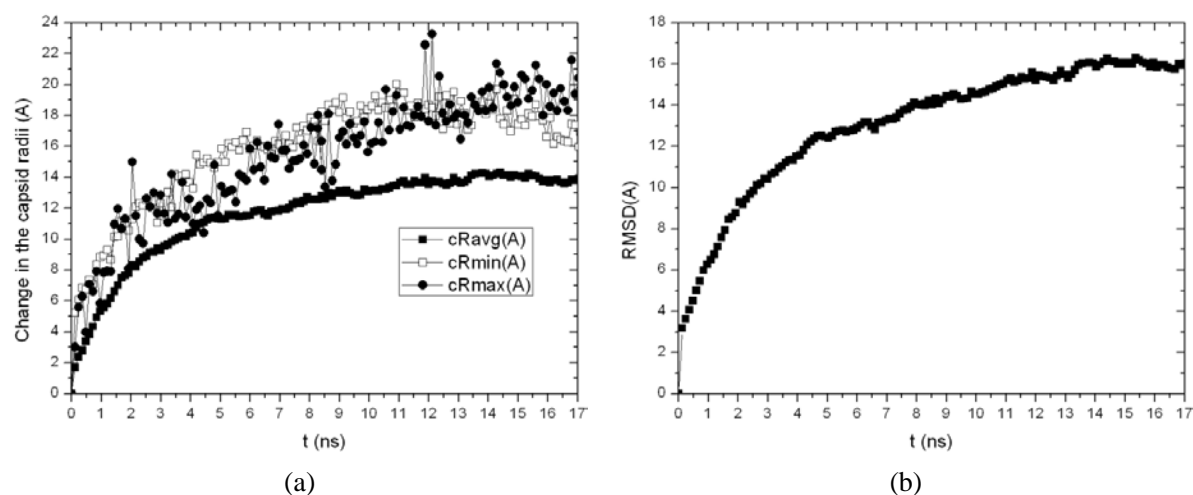




**Figure III.19** Analysis of 3ns MD/OPX trajectory of native CCMV capsid (Ncap): (a) time course of the change in capsid average radius from the starting configuration, (b) RMSD of atomic positions between snapshots of capsid backbone along the trajectory and the starting configuration, and (c) averaged thickness of water boundary surrounding the capsid backbone in X, Y and Z directions.

With the intermediate capsid structure (Icap) extracted from the above MD/OPX trajectory snapshot at 1.5ns, it was solvated in a larger water cube (25Å water boundary) for further simulation of the capsid swelling with 17.04ns dynamics captured. Analysis of the output MD/OPX trajectory shows that the capsid swells by 13.93Å in its backbone average radius

during 17.04ns with fluctuations found in the minimum and maximum radii of the capsid backbone (Figure III.20a). RMSD of atomic positions between capsid snapshots along the trajectory and its starting structure (Icap) increases from 0Å to 16.05Å in 17.04ns (Figure III.20b). The capsid was found to swell fast in early stage of the simulation with quick drop in the capsid average radius and sharp increase in the capsid RMSD, and the swelling slows down as the capsid approaches its near-equilibrium state.

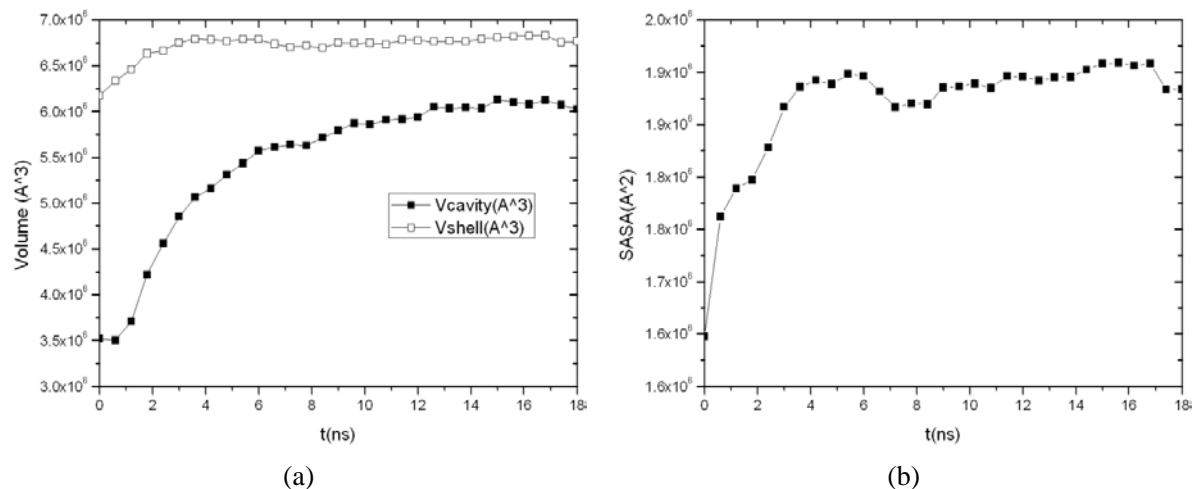


**Figure III.20** 17.04ns MD/OPX simulation of the swelling of capsid intermediate (Icap): (a) time course of the change in the capsid average radius and (b) RMSD of atomic positions between snapshots of capsid backbone along the trajectory and the initial structure.

Combining the above two MD/OPX trajectories into one, we obtained 18.54ns dynamics of CCMV capsid swelling from its native state. The total increase in the capsid average radius is 14.93Å. Analysis of the entire trajectory (Figure III.21) showed that the capsid cavity volume increases from  $3.52 \times 10^6 \text{ Å}^3$  to  $6.02 \times 10^6 \text{ Å}^3$  by 71% during the course of swelling. The volume of the capsid (i.e., shell) increases from  $6.17 \times 10^6 \text{ Å}^3$  to  $6.77 \times 10^6 \text{ Å}^3$  by ~10%. The capsid solvent-accessible surface area (SASA) was found to increase from  $1.65 \times 10^6 \text{ Å}^2$  to  $1.90 \times 10^6 \text{ Å}^2$  at 6ns and level off with small fluctuations thereafter. These indicate that regions of capsid



proteins that have been buried in the capsid interior in its native state are exposed to the solvent during capsid swelling, which also leads to increase in the capsid volume.

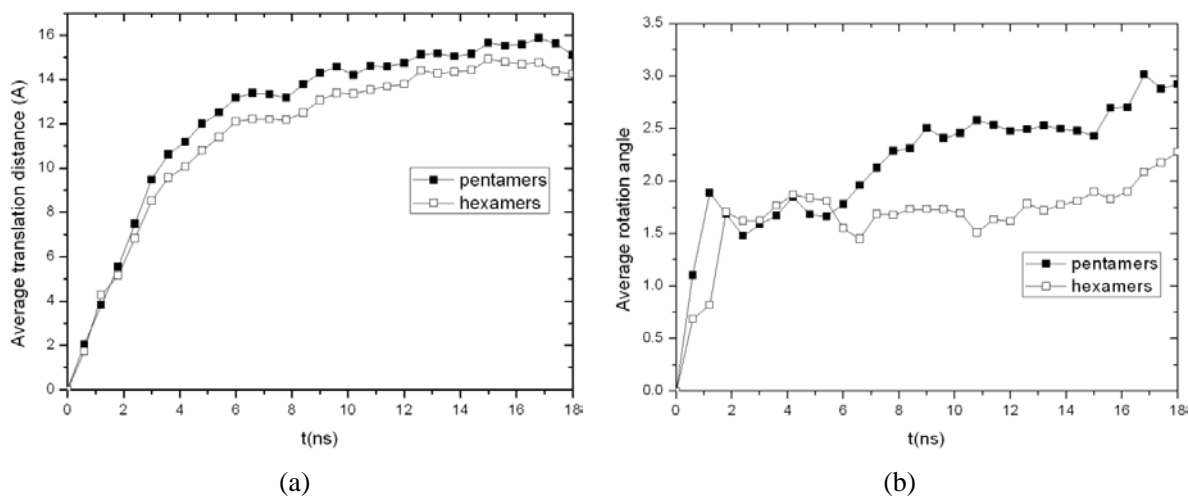


**Figure III.21** Analysis of CCMV capsid along the entire 18.54ns MD/OPX trajectory: (a) the volume of the capsid (i.e., protein shell) and its cavity, and (b) the capsid SASA.

### 9.3 CCMV Capsid Swelling Mechanisms

To characterize the collective motions of CCMV capsid during its swelling, we calculated the average COM translation distances and rotation angles of 12 pentamers and 20 hexamers along the trajectory obtained from above MD/OPX simulation. As shown in Figure III.22, pentamers were found to translate radially outwards by  $15.86\text{\AA}$  on average in 16.8ns and  $14.77\text{\AA}$  for hexamers, and there are slight drop at the end of the simulation; they also undergo counter-clockwise rotation by about  $2.92^\circ$  and  $2.28^\circ$  respectively with fluctuations along the trajectory. These agree with experimental data and other computational studies on the expansion of CCMV capsid from its native to the swollen state through rigid transformations of the capsomers [1, 18], even though the entire swelling process of CCMV capsid has not been

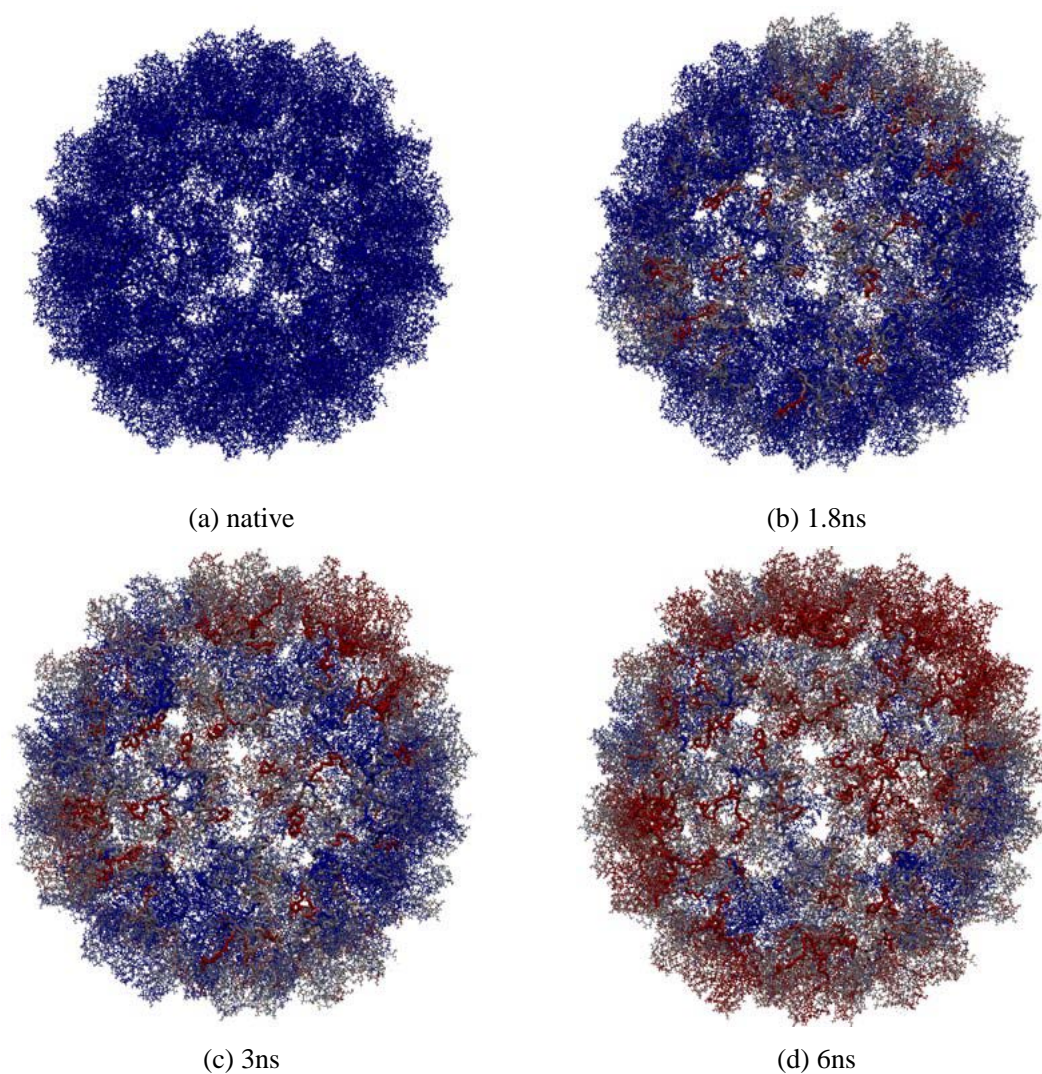
completely capture in 18.54MD/OPX simulation here. Therefore, the swelling of CCMV capsid in host media involves large-scale translation and rotation of the pentamers and hexamers, which was also found in the shrinkage of swollen CCMV capsid in vacuum studied earlier[13]. The capsomeres undergo cooperative motions through strongly-coupled allosteric interactions during viral capsid STs.

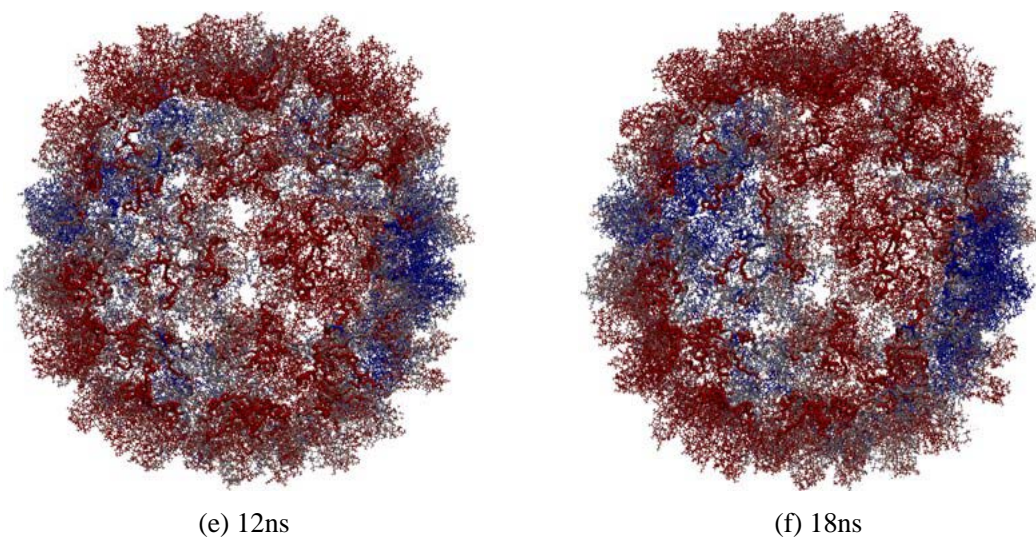


**Figure III.22** Collective motions of pentamers and hexamers during 18.54ns CCMV capsid swelling: time courses of (a) their average COM translation distances and (b) their average rotation angles calculated through structure fitting to the initial configurations in native CCMV capsid.

We now explore the swelling mechanisms of CCMV capsid by observing the atomistic motions of the capsid. Fig. 10 shows interior view of the capsid trajectory snapshots at (a) initial native state, (b) 1.8ns, (c) 3ns, (d) 6ns, (e) 12ns, and (f) 18ns. Atoms in the capsid are colored by their displacements from the original positions in a blue-white-red color scale (i.e., 0 Å for blue, 24 Å for red, and white as the midpoint) to investigate conformational changes in the capsid. N-terminal arms of the capsid proteins are highlighted in ribbons and they are

colored according to the displacements of atoms as well. It is seen that atomic displacements are not constrained to the icosahedral symmetry, i.e., atoms in the capsid do not change their colors simultaneously in a symmetry-preserving manner. Especially, the N-terminal arms change their color from blue to red during early stage of the simulation (see trajectory snapshots at 1.8ns and 3ns). This indicates that the N-terminal arms are highly dynamic and their structural changes trigger global expansion of the capsid.





**Figure III.23** Interior view of the capsid swelling trajectory snapshots with atoms colored by their displacements from the original positions in a BWR color scale (0Å for blue, 24Å for red, and white as the midpoint): (a) initial native state with all atoms in blue, (b) 1.8ns, (c) 3ns, (d) 6ns, (e) 12ns, and (f) 18ns. N-terminal arms of the capsid proteins are highlighted in ribbons and they are colored according to the displacements of atoms as well. It is seen that the N-terminal arms change their color from blue to red during early stage of the simulation and their structural changes trigger global expansion of the capsid. Swelling of CCMV capsid is a symmetry-breaking process involving local initiation and front propagation.

Similar to the shrinkage of swollen CCMV capsid in vacuum as simulated earlier[13], the ST of CCMV capsid during its swelling also involves “local initiation and front propagation”. The nucleation takes about 6ns to occur seen as the color change of two capsomers from blue to red in the trajectory snapshot. This is 14ns shorter than the nucleation time needed for the shrinkage of swollen CCMV capsid in vacuum. One reason is that N-terminal arms of capsid proteins were added in simulating the swelling of native CCMV capsid in host media and their fast fluctuations accelerate the nucleation of capsid ST. After the nucleation, the capsid ST then propagates across the capsid seen as spreading of the red areas. The propagation of cooperative motions from one capsomere to its neighboring ones takes place roughly on a timescale of 10ns. Thus the propagation speed can be calculated as ~0.6nm/ns with the average capsomere

distance considered as 6nm, which is similar to the ST in swollen CCMV capsid during its shrinkage in vacuum. Therefore, STs of viral capsids start with local conformational changes that break icosahedral symmetry of the capsid (e.g., fluctuations in N-terminal arms of capsid proteins), followed by global structural changes that involve large-scale cooperative translations and rotations of capsid structural units (e.g., capsomeres or protein subunits) through strong allosteric interactions.

#### **9.4 Dependence of CCMV Capsid Swelling on System Conditions**

The above simulations show that CCMV capsid swells over tens of nanoseconds, while such viral STs were observed to occur on timescales of milliseconds to minutes in experiments [51, 142, 143]. In the following, we present a list of simulations that indicate the swelling of CCMV capsid is largely dependent on system conditions. As listed in Table III.1, four simulations starting from the capsid native state were performed. For the first three simulations, the capsid was solvated in a water cube with two different water boundaries, i.e., 7Å (sim1) and 15Å (sim2 and sim3). For sim4, a truncated octahedron with 26Å water boundary was used for solvating the capsid (see a schematic representation in Fig. 3b). The water boundary here is the minimum distance between atoms on the capsid surface and the faces of water box (square face for cube and hexagonal face for truncated octahedron). Note that sim1 has been used to identify key elements for long-time simulation of the capsid swelling as discussed above. There are 522,000 atoms in CCMV capsid and the total number of atoms in the systems ranges from

~2.41 to ~2.86 million depending on the shape of water box and the thickness of water

boundary used for solvation.

**Table III.1** Simulations showing CCMV capsid swelling is largely dependent on system conditions.

Simulation	Shape of water box	Capsid interface <sup>a</sup>	Water boundary <sup>b</sup>	N <sub>atoms</sub> <sup>c</sup>	Ensemble <sup>d</sup>	Performance <sup>e</sup> (days/ns)	Time	Increase in radius <sup>f</sup>
sim1	Cube	2.4Å	7Å	2,411,918	NPT	2.60	3ns	1.35Å
sim2	Cube	2.4Å	15Å	2,845,915	NPT	2.95	3ns	1.58Å
sim3	Cube	2.4Å	15Å	2,845,915	NVT	2.36	5ns	10.48Å
sim4	Truncated octahedron	1.8Å	26Å	2,864,687	NPT	2.95	8ns	0.35Å

<sup>a</sup> Capsid interface is the minimum distance between water molecules and atoms on the capsid surface used for solvation, <sup>b</sup> water boundary is the minimum distance between atoms on the capsid surface and faces of the water box (square face for cube and hexagonal face for truncated octahedron); <sup>c</sup> there are 522,000 atoms in CCMV capsid and the total number of atoms in the systems ranges from ~2.41 to ~2.86 million depending on the shape of water box used for solvation and the thickness of water boundary; <sup>d</sup> NPT is constant pressure (1atm) and temperature (298.15K) ensemble and NVT is constant volume (unchanged from the initial dimensions of systems prepared with VMD) and temperature (298.15K) ensemble; <sup>e</sup> performance here is for running simulations with 256 processors on Indiana University Big Red cluster; and <sup>f</sup> the capsid radius calculated is the average distance of non-hydrogen atoms in the capsid backbone to their center of mass.

For MD simulations, periodic boundary conditions were applied. Constant volume and temperature control (i.e., NVT ensemble) was used for sim3, and the others were performed under constant pressure and temperature control (i.e., NPT ensemble). The specified temperature was 298.15K and the pressure was 1atm for NPT simulations. With the above settings, NAMD was used to simulate these four systems to investigate factors of CCMV capsid swelling. It was found that the unit cell for systems prepared with VMD “solvate” plugin by



using its default parameters need to be compressed in order to reach 1atm pressure in NPT simulations (e.g., sim1 and sim2). Without pressure control, the systems would have a pressure much lower than 1atm in NVT simulation (e.g., sim3). The analysis of the output trajectories is as follows.

CCMV capsid was found to swell much faster and more dramatically in sim3 under NVT control (i.e., constant volume, low pressure) than in NPT simulations. While the capsid swells only 1.58Å in its average radius during 3ns NPT simulation in sim2 and 1.35Å in sim1, the capsid as simulated under NVT control swells 10.03Å in 3ns and 10.48Å in 5ns in sim3. During the 5ns significant swelling, the capsid thickens by 16.91Å with its exclusion volume increased from  $6.17 \times 10^6 \text{ Å}^3$  to  $6.89 \times 10^6 \text{ Å}^3$  and the capsid SASA increased from  $1.65 \times 10^6 \text{ Å}^2$  to  $1.92 \times 10^6 \text{ Å}^2$ ; the volume of capsid cavity is increased by 42% from  $3.52 \times 10^6 \text{ Å}^3$  to  $5.01 \times 10^6 \text{ Å}^3$ . In early stage of the simulation, water molecules flux from the capsid cavity into the capsid to solvate proteins and small regions of vacuum are formed inside the capsid. As the capsid continues swelling, water molecules flux from the capsid exterior into the capsid cavity. The density of water inside the capsid cavity decreases significantly from  $975.51 \text{ kg/m}^3$  to  $752.83 \text{ kg/m}^3$  at the end of the 5ns simulation, while density of water outside of the capsid stays roughly constant.

When pressure control is applied as in sim1 and sim2, the system is compressed to reach 1atm pressure and the swelling rate of CCMV capsid is decreased. The capsid cavity volume, the exclusion volume of the capsid and its SASA are also found to increase accompanying the capsid swelling. The density of water in the capsid cavity decreases during early stage of the simulation due to swelling of the capsid and the flux of water molecules into the capsid to

solvate proteins. It then increases as the capsid swelling slows down and the water molecules flux from the capsid exterior into the capsid cavity under the pressure control. In the case of sim2, the water density inside the capsid cavity decreases from  $975.48\text{kg/m}^3$  to  $797.32\text{kg/m}^3$  at 1ns and then increases to  $821.25\text{kg/m}^3$  at 3ns.

To understand the different swelling behaviors of CCMV capsid found in the NPT and NVT simulations, we compare the system pressure profiles of sim2 and sim3. The pressure of the entire NVT simulated system in sim2 was found to level off around -178atm after equilibration, which is much lower than that of the NPT equilibrated system in sim3, i.e., ~1atm. The high negative pressure was encountered in sim2 because the system volume was kept constant and unphysically large. Next, with the simulation box divided into horizontal slices of  $\sim 10\text{\AA}$  thickness in Z-direction, we calculate the local pressure of each slice by using the pressure profile analysis module provided in NAMD [75, 144]. Calculation results showed that the average pressure difference between slices that sandwich CCMV capsid at the bottom and the top (both from the capsid cavity surface to the capsid exterior) is 88atm at 0.1ns in sim3, which is ~13atm greater than in sim2. This suggests the pressure gradient across the capsid in sim3 is higher than that in the latter. As the capsid swells, the pressure of the system slice right below or above the capsid (i.e., pressure of water outside of the capsid) keeps increasing. It takes 0.6ns to decrease the pressure gradient across the capsid to zero in sim2, while 1ns (i.e., 0.4ns longer) in sim3. As a result, the capsid swells faster and more dramatically in NVT simulation sim3 than in NPT simulation sim2.



It was seen that water molecules flux into the capsid to solvate proteins during early stage of the above simulations and small regions of vacuum can be formed inside the capsid. This was also discovered in the MD simulation of satellite tobacco mosaic virus [52]. It could result from a large parameter in system preparation defined for the distance between water molecules and the capsid surface. In the VMD “solvate” plugin, the default value for the minimum distance between water molecules and the solute is 2.4Å, while it can be as small as 1.8Å as obtained through our NPT simulations (sim1 and sim2). In light of this finding, we used 1.8Å as the minimum distance between water molecules and atoms on the capsid surface for solvating the capsid in sim4. With this, the flux of water molecules into the capsid proteins is greatly reduced. CCMV capsid was found to hardly swell with its average radius increased by only 0.35Å during the 8ns simulation. The capsid cavity volume actually decreases from  $3.52 \times 10^6 \text{ Å}^3$  to  $3.24 \times 10^6 \text{ Å}^3$  as the capsid increases its exclusion volume, i.e., regions of the capsid proteins protrude into the cavity. The density of water in the capsid cavity was found to increase from 974.41 kg/m<sup>3</sup> to 1035.69 kg/m<sup>3</sup> with water molecules still flux from the capsid interior into the capsid cavity. The water density outside of the capsid stays constant (~990kg/m<sup>3</sup>) close to that obtained for pure water (997.05 kg/m<sup>3</sup>) at 298.15K and 1atm pressure. Therefore, the high density of water surrounding the capsid compared with that of the above three simulations and the incompressibility of water stop the capsid from large-scale swelling.

Therefore, the swelling of CCMV capsid in a host medium is strongly dependent on simulation parameters (e.g., the interfacial distance between water molecules and capsid

proteins used for solvation and system water boundary) and physical factors (e.g., the density of water inside the capsid cavity and that outside of the capsid, and the pressure gradient across the capsid). The capsid swells slightly over a timescale of nanoseconds when the density of water surrounding and within the capsid is close to that of pure water (sim4). While this might share the same timescale for capsid swelling as observed from experiments (i.e., milliseconds to minutes) with the real system conditions reflected, it is not feasible to simulate the entire capsid swelling process with the current computational power[5]. Under specific system conditions, one can simulate the swelling of CCMV capsid over tens of nanoseconds, which provides insights into its ST pathways.

In addition to factors discussed above, discrepancies between the timescale for capsid swelling obtained from our simulations and that observed from experiments include the following. (1) The free energy landscape is populated by many minima. Thus along the pathway connecting the original to the final structure the system can reside for appreciable time at these minima. In the transition for one minimum to the next, many transition states and associated energy barriers are encountered. (2) The free energy landscape can be so populated with minima that the system is driven by a higher-order entropy, i.e., the system loses track of where it is going although it is driven by an overall energy gradient; this is a complex analog of a two well potential with a broad, shallow well and a narrow shallow one; in that case the system could prefer to be in the relatively high energy well. (3) The dynamics of viral capsid is likely highly frictional; thus despite a large free-energy gradient the kinetics can be slow. (4) Factors enhancing attractive forces between capsomeres (e.g., divalent cations) will repress capsid

swelling, which requires overcoming the energy barriers.

## 9.5 Prospects of multiscale MD/OPX simulation

The pH-induced swelling of CCMV has been suggested to be a first order phase transition at low salinity ( $I=0.2\text{M}$ ) with hysteresis found in its titration curve [126]. However, divalent cations, such as  $\text{Mg}^{2+}$  and  $\text{Ca}^{2+}$ , are known to bind CCMV capsid proteins at the quasi-threefold axes, and thus stabilize the capsid and the complete virus [18, 39]. By adding magnesium ions to the solution ( $[\text{Mg}^{2+}]=0.01\text{M}$ ), hysteresis in the titration curve of CCMV can be abolished [40, 126] and thus CCMV swelling becomes fully reversible. This indicates that virus swelling can become a second order phase transition in the presence of divalent cations. It was also found that the effect of a high concentration in potassium ions (e.g.,  $0.8\text{M KCl}$ ) is identical to that of magnesium for eliminating hysteresis in the CCMV titration curve. Furthermore, as salinity is increased from  $0.2\text{M}$  to  $0.3\text{M}$ , the hysteresis loop size is reduced by a factor of 2. While the above controlling factors for the swelling of complete CCMV have been determined, those for the empty CCMV capsid swelling are still under investigation [145].

Our all-atom multiscale simulation can be used to create a phase diagram for CCMV capsid mapping out regions in the space of pH, salinity, concentration of divalent cations, and temperature wherein first and second order transitions take place. As capsid fluctuations are believed to play an important role in viral STs, identifying conditions where critical fluctuations are realized will be of great interest. The simulations will be used to suggest the experimental

conditions wherein large fluctuations and the most dramatic STs occur. From our simulations, swelling of CCMV capsid was found to start locally and then spread across the capsid. However, one might suggest that viral STs occur simultaneously. If this is the case, there should be instantaneous communication of proteins across the viral capsid, e.g., as follows from Coulomb interactions. Therefore, decreasing salinity will increase the length scale of effective interactions, which is limited to the Debye length. In this way, we will determine theoretically if there is an observable transition between the nucleation/propagation and the global/simultaneous ST mechanism as salinity changes. This suggests that at low salinity the system should be highly cooperative and global so that critical fluctuations would be repressed; alternatively, the high salinity system is expected to act more like a two or three dimensional system and display critical behavior.

As found through our MD/OPX simulation, the N-terminal arms of CCMV capsid proteins are highly dynamic. They undergo fast fluctuations during early stage of the simulation and their structural changes trigger global expansion of the capsid. Such fluctuations are found to take place in other viral capsids. They are believed to play important roles in packaging viral genome during virus maturation or engineered nanomaterial synthesis by using viral capsids as molecular containers [24]. For example, the N-terminus of the poliovirus capsid protein VP1 that is entirely internal in native virion becomes externalized upon cell attachment, and the exposed N-terminus of VP1 was shown to be responsible for liposome binding [43, 44]. These fluctuations can be predicted from our simulations and used to guide experimental studies, e.g., chemical labeling[146]. If the protein terminal domains are protruding/retracting in a

fluctuating manner, then the degree of labeling will increase when such fluctuations are large.

Furthermore, the depth of penetration of the labeling should increase as critical fluctuation intensity increases.

## Part IV. Summary and Prospects

Viruses and other nanoparticles display mixed microscopic/macroscale characteristics. It is natural to develop an understanding of their dynamics via a multiscale analysis of the Liouville equation following prescriptions introduced for the study of Brownian motion. However the internal dynamics of the atoms constituting a nanoparticle introduces conceptual and technical difficulties associated with a description involving both the atomistic and nanometer scale properties of these systems and the potential overcounting of degrees of freedom. To overcome these difficulties we introduce a “nanocanonical ensemble” method to facilitate the all-atom multiscale analysis (AMA) of the  $N$ -atom Liouville equation. Slow variables (i.e., order parameters) are identified to capture the internal structure of nanoparticles, but not to be thought of as new dynamical variables. Rather they are a way to make the multiscale character of the  $N$ -atom probability density explicit. Using this formulation and the chain rule, the need for tedious bookkeeping to ensure conservation of the total number of degrees of freedom was avoided.

A further advance of our AMA approach is to ensure removal of all secular behavior in the  $N$ -atom probability density and not just that of a reduced distribution. It is straightforward to remove the secular behavior in the  $N$ -atom probability density for structureless nanoparticles by integration over a reduced set of atomic degrees of freedom (i.e. of the host atoms), but it cannot be readily carried out for the structured nanoparticles. In AMA, the fast variable

integration is avoided by using the fundamental hypothesis of statistical mechanics, “the long-time and ensemble averages are equivalent near equilibrium for systems with slowly varying collective behaviors”. The nanocanonical formulation and integration-free removal of secular overcome conceptual and technical difficulties encountered in a multiscale analysis of the Liouville equation for structured nanoparticles, which leads to a stochastic (FP or Smoluchowski) equation yielding the coarse-grained evolution of the slow variables. Being based on a calibrated interatomic force field, our method can yield parameter-free universal models for nanoparticle dynamics.

Following from the AMA formulation, order parameters capturing viral dilatational structural transitions (STs) are constructed, a FP equation for their long timescale dynamics is derived, and a transition state ansatz is presented for obtaining approximate solutions to the FP equation in the inertial limit. While viral free energy and friction coefficients need to be computed from a calibrated interatomic force field and the FP equation solved to realize the full implications of this approach, we calculate the free energy barrier for the ST of NoV capsid from its procapsid form to the compact capsid form based on crude estimates. Calculation results show that the transition has a lower free energy barrier than that for the transition of native poliovirus to a cell entry intermediate, and it depends on how we change the pH to induce the ST: when pH is lowered from 7.6 to 5.0, it occurs on 0.1 sec timescale with a free energy barrier of 10.86 kcal/mol; with a less drastic pH change (pH lowered to 5.8 or 5.5 instead of 5.0), the ST occurs more slowly on a 100 sec timescale with a higher free energy barrier of 14.95 kcal/mol. This transition state estimate of the free energy barrier provides an

upper bound for the capsid ST, i.e. the single order parameter model likely misses lower energy barrier pathways for the capsid transition, and more order parameters that characterize the deformation of viruses apart from their COM position, orientation, dilatational order parameter and their associated momenta are needed.

To capture the large-scale low-frequency coherent motions of nanosystems, we construct collective modes and order parameters with basis functions (e.g., orthogonal polynomials or harmonic functions) of atomic coordinates of a reference configuration or atomic displacements between two known configurations. A structural change in the virus is expressed as a summation of the chosen collective modes weighted by order parameters that are determined to minimize the root mean square residuals due to mode truncation. These collective modes are demonstrated to be capable of capturing nanoscale dynamics of capsomers during the expansion of native CCMV capsid.

Based on AMA and the slowly-varying characteristics of order parameters, a multiscale MD/OPX computational modeling approach is developed to simulate large bionanosystems, especially viruses of our interest. In the implementation, replica short MD runs with random atomic velocity initializations are implemented to estimate the ensemble average rate of change in the order parameters, which is then used to extrapolate the state of the system over a time period that is much longer than the  $10^{-14}$ -second timescale of fast atomic vibrations and collisions. MD/OPX essentially solves the Langevin equations for the stochastic dynamics of order parameters implicitly. Since the timescale for the evolution of order parameters is much larger than that of atomistic vibrations and collisions, the OP extrapolation time (e.g., 60ps) can



be many orders of magnitude greater than the MD simulation timestep. Traditional coarse-grained models start with lumping clusters of atoms into structureless beads and simulate their time evolution, which misses the feedback between the atomistic fluctuations and the coherent nanoscale dynamics. They ignore the facts that system coherent motions involve all the atoms moving collectively. In contrast, our MD/OPX captures the cross-talk between the order parameters and atomistic variables rigorously and addresses the atomistic fluctuations and nanoscale coherent motions simultaneously.

MD/OPX is validated on the STs of CCMV capsid. Simulations show that the swollen state of the capsid undergoes significant shrinkage in vacuum and the MD/OPX trajectory analysis results agree well with those of direct MD simulation. MD/OPX is found to be 9.76 times faster than NAMD. After the validation, MD/OPX is applied to investigating the ST mechanisms of viral capsids still by choosing CCMV as our model system. Insights into the stabilities of the native and swollen states of CCMV capsid in vacuum were obtained through short-time MD simulation. With the results showing that native CCMV capsid becomes stable after short, small shrinkage in 10ns, while its swollen state shrinks over a timescale that is much longer than 10ns, long-time MD/OPX simulation was run on the swollen state with 200ns dynamics of its shrinkage captured. Shrinkage of swollen CCMV capsid was found to be an energy-driven, symmetry-breaking process that involves large-scale translation and rotation of pentamers and hexamers in the capsid. The capsomeres undergo cooperative motions through strongly-coupled allosteric interactions during capsid shrinkage.

While native CCMV capsid is fairly stable in vacuum, it was found to swell in a host medium

at pH 7.0 and ionic strength  $I=0.2M$  as obtained through short-time MD simulations. The application of MD/OPX to the swelling of native CCMV capsid also shows that the N-terminal arms of capsid protein subunits are highly dynamic and their fast fluctuations trigger global expansion of the capsid. The capsid swelling is also a symmetry-breaking process involving local initiation and front propagation. Viral STs start locally and then propagate across the capsid, i.e., they proceed via intermediate states that are not constrained to the capsid symmetry. This suggests that it is not appropriate to use symmetry-constrained models to study the pathways and mechanisms of viral STs.

With our AMA method and MD/OPX, viral dynamics are predicted from first principles. MD/OPX accelerates MD for long-time simulation of viruses, as well as other bionanosystems, and it addresses rapid atomistic fluctuations and slowly-varying nanoscale dynamics simultaneously. However, the following factors limit the applicability of MD/OPX as presently implemented: (1) structural changes that involve bond breaking (e.g., proteolysis of capsid proteins during virus maturation [10, 48, 147]) necessitate the incorporation of a quantum chemistry model, i.e., quantum mechanics/molecular mechanics (QM/MM), or a force field that can account for bond breaking, (2) highly fluctuating OPs that do not display linear trend lines are not appropriate for extrapolation; therefore inclusion of additional OPs does not necessarily improved accuracy or performance, (3) systems that involve fast inter-diffusion of small molecules and host particles cannot necessarily be described using OPs as presently implemented MD/OPX; to simulate such systems, one must use OP density variables [102, 148] and implement a module analogous to “resolution” explained in Section 5.3, and (4) inherent

limitations in the CHARMM and AMBER force fields (e.g., dielectric channeling) will also be part of MD/OPX predictions as presently implemented; however, as NAMD is improved, so will MD/OPX.

In the above MD/OPX simulation of CCMV capsid swelling in a host medium, water molecules and ions are modeled explicitly. Instead, an implicit solvent approach can be used to accelerate the simulations. Our AMA formulation also justifies the development of a direct Langevin solver for stochastic dynamics of the order parameters. 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  $\sim 0.5$  million atoms, many systems of biological interest are supra-million atoms in size (e.g. large viruses and cell membranes) and as the size of the simulated system increases, the simulation speed-up of our multiscale approach over direct MD will become greater.

Viral dynamics have been captured over timescales of tens to hundreds of nanoseconds with our multiscale MD/OPX simulations. While they are still short compared with the millisecond-to-minute timescales for viral STs believed to take place in experiments [51, 142, 143], MD/OPX and the direct Langevin solver of system order parameters [11, 102] based on our AMA methodology hold great promise for long-time simulation of viruses and other bionanosystems. To do so, one must address the possibility of large activation energy for barrier

crossing and additional extremely slow order parameters, such as the  $H^+$  concentration. The latter would be important, e.g., if the long-timescale of viral STs is associated with the diffusion of protons to eliminate supra-micron scale non-uniformities accompanying addition of  $H^+$ . 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.

## Appendix A: Decomposition of the Liouville equation

The Liouville equation for the  $N$ -atom probability density  $\rho(\bar{p}_1, \bar{r}_1, \dots, \bar{p}_N, \bar{r}_N, t)$  is taken to consist of a nanoparticle of  $N^*$  atoms and a host medium of  $N - N^*$  atoms. The Liouville equation for this system can be written as

$$\frac{\partial \rho}{\partial t} = - \sum_{i=1}^N \left\{ \frac{\bar{p}_i}{m_i} \cdot \frac{\partial}{\partial \bar{r}_i} + \bar{F}_i \cdot \frac{\partial}{\partial \bar{p}_i} \right\} \rho \equiv \mathcal{L} \rho, \quad (\text{A.1})$$

where  $\bar{F}_i = - \left( \frac{\partial V}{\partial \bar{r}_i} \right)_{\bar{r}_{j \neq i}}$  is the force on atom  $i$ .

Our development starts by considering  $\rho$  to be dependent on the nanoparticle COM position  $\bar{r}^*$  and momentum  $\bar{p}^*$ , as well as individual atomic positions and momenta. Nanoparticle mass  $m^*$  and COM variables are expressed in terms of atomic variables via

$$\begin{Bmatrix} \bar{r}^* \\ \bar{p}^* \\ m^* \end{Bmatrix} = \sum_{i=1}^N \begin{Bmatrix} m_i \bar{r}_i / m^* \\ \bar{p}_i \\ m_i \end{Bmatrix} \Theta_i, \quad (\text{A.2})$$

where  $\Theta_i = 1$  (atom  $i$  in the nanoparticle),  $= 0$  (otherwise). In the multiscale analysis with slow variables  $\bar{p}^*$  and  $\bar{r}^*$ , the  $N$ -body probability density is taken to have the dependence  $\rho^*(\bar{p}_1, \bar{r}_1, \dots, \bar{p}_N, \bar{r}_N, \bar{p}^*, \bar{r}^*, t)$ . Upon invoking the chain rule,  $\rho^*$  is found to satisfy

$$\frac{\partial \rho^*}{\partial t} = \mathcal{L}^{b*} \rho^* - \left[ \frac{\bar{p}^*}{m^*} \cdot \frac{\partial}{\partial \bar{r}^*} + \sum_{i=1}^N \Theta_i \bar{F}_i \cdot \frac{\partial}{\partial \bar{p}^*} \right] \rho^*, \quad (\text{A.3})$$

where a superscript  $*$  on an operator implies that  $\bar{p}_1, \bar{r}_1, \dots, \bar{p}_N, \bar{r}_N$ ,  $\bar{r}^*$  and  $\bar{p}^*$  are the

independent variables. Hence  $\left(\frac{\partial}{\partial \bar{r}_i}\right)^*$  implies an  $\bar{r}_i$  partial derivative keeping  $\bar{r}_{j \neq i}$ , all  $\bar{p}_j$ ,  $\bar{r}^*$  and  $\bar{p}^*$  constant. Finally,

$$\mathcal{L}^{b*} = - \sum_{i=1}^N \left\{ \frac{\bar{p}_i}{m_i} \cdot \frac{\partial}{\partial \bar{r}_i} + \bar{F}_i \cdot \frac{\partial}{\partial \bar{p}_i} \right\}^* . \quad (\text{A.4})$$

This concludes the first reformulation.

Next we introduce a scaling ansatz and further definitions. Taking the scaled variables  $m^* = \varepsilon^{-2} m$ ,  $\bar{p}^* = \varepsilon^{-1} \bar{P}$  and  $\bar{r}^* = \varepsilon^{-1} \bar{R}$ , and introducing a set of scaled times  $t_n = \varepsilon^{2n} t, n=0,1,\dots$ , the  $N$ -atom probability density is taken to have the dependence  $\rho^{**}(\bar{p}_1, \bar{r}_1, \dots, \bar{p}_N, \bar{r}_N, \bar{P}, \bar{R}; t_0, t_1, \dots; \varepsilon)$ . With this and the chain rule  $\rho^{**}$  satisfies

$$\sum_{n=0} \varepsilon^{2n} \left( \frac{\partial}{\partial t_n} - \mathcal{L}_n \right) \rho^{**} = 0. \quad (\text{A.5})$$

Letting a superscript  $**$  on an operator indicates that  $\bar{p}_1, \bar{r}_1, \dots, \bar{p}_N, \bar{r}_N$ ,  $\bar{P}$  and  $\bar{R}$  are the independent variables, we find

$$\mathcal{L}_0 = - \sum_{i=1}^N \left\{ \frac{\bar{p}_i}{m_i} \cdot \frac{\partial}{\partial \bar{r}_i} + \bar{F}_i \cdot \frac{\partial}{\partial \bar{p}_i} \right\}^{**}, \quad (\text{A.6})$$

$$\mathcal{L}_1 = - \left[ \frac{\bar{P}}{m} \cdot \frac{\partial}{\partial \bar{R}} \right]^{**} - \left[ \bar{f} \cdot \frac{\partial}{\partial \bar{P}} \right]^{**}, \quad (\text{A.7})$$

where  $\varepsilon \bar{f} = \bar{F}_1 + \dots + \bar{F}_{N^*}$  is the net force on the virus and  $\mathcal{L}_n = 0$  for  $n \geq 2$ . The  $**$  will be dropped for simplicity henceforth.

While above formulation shows the decomposition of the  $N$ -atom Liouville equation with slow variables  $\bar{P}$  and  $\bar{R}$ , it can be generalized with a set of slow variables. In the case of

viral dilatational STs with slow variable set  $\{ \bar{P}, \bar{R}, \Pi, \Phi \}$ , the scaled Liouville operators have the following form:

$$\mathcal{L}_0 = - \sum_{i=1}^N \left[ \frac{\bar{p}_i}{m_i} \cdot \frac{\partial}{\partial \bar{r}_i} + \bar{F}_i \cdot \frac{\partial}{\partial \bar{p}_i} \right], \quad (\text{A.8})$$

$$\mathcal{L}_1 = - \frac{\bar{P}}{m} \cdot \frac{\partial}{\partial \bar{R}} - \bar{f} \cdot \frac{\partial}{\partial \bar{P}} - \frac{\Pi}{m'} \frac{\partial}{\partial \Phi} - g \frac{\partial}{\partial \Pi}. \quad (\text{A.9})$$

## Appendix B: Fourier/bilateral Laplace transformation

Given

$$\hat{f}(k) = \int_{-\infty}^{+\infty} dx e^{kx} f(x) \quad (\text{B.1})$$

one may show that the inverse relation reads

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} dk e^{-ikx} \hat{f}(ik) \quad (\text{B.2})$$

with  $i = \sqrt{-1}$ . For two functions  $f_1(x)$  and  $f_2(x)$  one has

$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} dk e^{-ikx} \hat{f}_1(ik) f_2(ik) = \int_{-\infty}^{+\infty} dx' f_1(x-x') f_2(x'). \quad (\text{B.3})$$

Results in Sect. III follow upon generalization to six dimensions. For example, one may obtain

the inversion formula

$$\frac{\Psi(\bar{\mu}, \bar{\kappa}, \underline{t})}{\Xi(\beta, \bar{\mu}, \bar{\kappa})} = \frac{1}{(2\pi)^6} \int d^3 p d^3 r e^{-i(\bar{\mu} \cdot \bar{p} + \bar{\kappa} \cdot \bar{r})} \frac{W(i\bar{p}, i\bar{r}, \underline{t})}{Q(\beta, i\bar{p}, i\bar{r})} \quad (\text{B.4})$$

Thus  $\Psi/\Xi$  is related to the integration of  $W/Q$  over all imaginary momenta and positions.



## Bibliography

1. Liu, H.J., et al., *Pseudo-atomic models of swollen CCMV from cryo-electron microscopy data*. Journal of Structural Biology, 2003. **142**(3): p. 356-363.
2. Flint, S.J., et al., *Principles of Virology: Molecular Biology, Pathogenesis, and Control*, 3rd ed. ASM Press, 2009.
3. Shepherd, C.M., et al., *VIPERdb: a relational database for structural virology*. Nucleic Acids Research, 2006. **34**: p. D386-D389.
4. Johnson, J.E. and W. Chiu, *Structures of virus and virus-like particles*. Curr Opin Struct Biol, 2000. **10**(2): p. 229-35.
5. Klepeis, J.L., et al., *Long-timescale molecular dynamics simulations of protein structure and function*. Curr Opin Struct Biol, 2009. **19**(2): p. 120-7.
6. Angelescu, D.G. and P. Linse, *Modelling of icosahedral viruses*. Current Opinion in Colloid & Interface Science, 2008. **13**(6): p. 389-394.
7. Hogle, J.M., *Poliovirus cell entry: Common structural themes in viral cell entry pathways*. Annual Review of Microbiology, 2002(56): p. 677-702.
8. Belnap, D.M., et al., *Three dimensional structure of poliovirus receptor bound to poliovirus*. Proc. Natl. Acad. Sci., 2000. **97**: p. 73-78.
9. Miao, Y. and P.J. Ortoleva, *Viral structural transitions: an all-atom multiscale theory*. Journal of Chemical Physics, 2006. **125**(21): p. 214901.
10. Canady, M.A., H. Tsuruta, and J.E. Johnson, *Analysis of Rapid, Large-scale Protein Quaternary Structural Changes: Time-resolved X-ray Solution Scattering of Nudaurelia capensis  $\omega$  Virus (NoV) Maturation*. Journal of Molecular Biology, 2001(311): p. 803-814.
11. Miao, Y. and P. Ortoleva, *All-atom multiscale and new ensembles for dynamical nanoparticles*. Journal of Chemical Physics, 2006. **125**(4): p. 44901.
12. Miao, Y. and P.J. Ortoleva, *Molecular dynamics/order parameter extrapolation for bionanosystem simulations*. J Comput Chem, 2008. **30**(3): p. 423-437.
13. Miao, Y. and P. Ortoleva, *Viral Structural Transition Mechanisms Revealed via Multiscale Molecular Dynamics/Order Parameter eXtrapolation Simulation*. Biopolymers, In press, 2009.
14. Miao, Y., J. Johnson, and P. Ortoleva, *All-atom Multiscale Simulation of CCMV Capsid Swelling*. J Phys Chem B, 2009. **submitted**.
15. Johnson, J.E. and J.A. Speir, *Quasi-equivalent viruses: A paradigm for protein assemblies*. Journal of Molecular Biology, 1997. **269**(5): p. 665-675.
16. Caspar, D.L.D. and A. Klug, *Physical Principles in Construction of Regular Viruses*. Cold Spring Harbor Symposia on Quantitative Biology, 1962. **27**: p. 1-24.
17. Johnson, J.E., *Functional implications of protein-protein interactions in icosahedral viruses*. Proceedings of the National Academy of Sciences of the United States of

- America, 1996. **93**(1): p. 27-33.
18. Speir, J.A., et al., *Structures of the Native and Swollen Forms of Cowpea Chlorotic Mottle Virus Determined by X-Ray Crystallography and Cryoelectron Microscopy*. Structure, 1995. **3**(1): p. 63-78.
  19. Hogle, J.M., M. Chow, and D.J. Filman, *Three-dimensional structure of poliovirus at 2.9 Å resolution*. Science, 1985. **229**(4720): p. 1358-65.
  20. Bink, H.H.J. and C.W.A. Pleij, *RNA-protein interactions in spherical viruses*. Archives of Virology, 2002. **147**(12): p. 2261-2279.
  21. Douglas, T. and M. Young, *Virus particles as templates for materials synthesis*. Advanced Material, 1999. **11**(8): p. 679.
  22. Endy, D. and J. Yin, *Toward antiviral strategies that resist viral escape*. Antimicrobial Agents and Chemotherapy, 2000. **44**(4): p. 1097-1099.
  23. Olson, A.J., Y.H.E. Hu, and E. Keinan, *Chemical mimicry of viral capsid self-assembly*. Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(52): p. 20731-20736.
  24. Douglas, T., et al., *Protein Engineering of a Viral Cage for Constrained Nanomaterials Synthesis*. Advanced Material, 2002. **14**(6): p. 415-418.
  25. Chen, C., et al., *Nanoparticle-templated assembly of viral protein cages*. Nano Letters, 2006. **6**(4): p. 611-615.
  26. Huang, X.L., et al., *Self-assembled virus-like particles with magnetic cores*. Nano Letters, 2007. **7**(8): p. 2407-2416.
  27. Dixit, S.K., et al., *Quantum dot encapsulation in viral capsids*. Nano Letters, 2006. **6**(9): p. 1993-1999.
  28. Zandi, R., et al., *Origin of icosahedral symmetry in viruses*. Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(44): p. 15556-15560.
  29. Bruinsma, R.F., et al., *Viral self-assembly as a thermodynamic process*. Physical Review Letters, 2003. **90**(24).
  30. Ceres, P. and A. Zlotnick, *Weak protein-protein interactions are sufficient to drive assembly of hepatitis B virus capsids*. Biochemistry, 2002. **41**(39): p. 11525-31.
  31. Zlotnick, A., *Are weak protein-protein interactions the general rule in capsid assembly?* Virology, 2003. **315**(2): p. 269-74.
  32. Zlotnick, A., *Theoretical aspects of virus capsid assembly*. Journal of Molecular Recognition, 2005. **18**(6): p. 479-490.
  33. Zlotnick, A. and S.J. Stray, *How does your virus grow? Understanding and interfering with virus assembly*. Trends in Biotechnology, 2003. **21**(12): p. 536-542.
  34. Zlotnick, A., et al., *Mechanism of capsid assembly for an icosahedral plant virus*. Virology, 2000. **277**(2): p. 450-456.
  35. Zlotnick, A., et al., *A theoretical model successfully identifies features of hepatitis B virus capsid assembly*. Biochemistry, 1999. **38**(44): p. 14644-14652.
  36. McPherson, A., *Micelle formation and crystallization as paradigms for virus assembly*. Bioessays, 2005. **27**(4): p. 447-58.

37. Endres, D., et al., *A reaction landscape identifies the intermediates critical for self-assembly of virus capsids and other polyhedral structures*. Protein Science, 2005. **14**(6): p. 1518-1525.
38. Pankavich, S., et al., *Self-assembly of nanocomponents into composite structures: Derivation and simulation of Langevin equations*. The Journal of Chemical Physics, 2009. **130**(19): p. 194115.
39. Liepold, L.O., et al., *Structural transitions in Cowpea chlorotic mottle virus (CCMV)*. Phys. Biol., 2005(2): p. S166–S172.
40. Adolph, K.W., *Structural Transitions of Cowpea Chlorotic Mottle Virus*. J. gen. Virology, 1975(28): p. 147-154.
41. Taylor, D.J., et al., *Large-Scale, pH-Dependent, Quaternary Structure Changes in an RNA Virus Capsid Are Reversible in the Absence of Subunit Autoproteolysis*. Journal of Virology, 2002. **76**(19): p. 9972-9980.
42. Wikoff, W.R., et al., *Time-resolved molecular dynamics of bacteriophage HK97 capsid maturation interpreted by electron cryo-microscopy and X-ray crystallography*. Journal of Structural Biology, 2006: p. 300-306.
43. Fricks, C.E. and J.M. Hogle, *Cell-Induced Conformational Change in Poliovirus - Externalization of the Amino Terminus of Vp1 Is Responsible for Liposome Binding*. Journal of Virology, 1990. **64**(5): p. 1934-1945.
44. Belnap, D.M., et al., *Molecular tectonic model of virus structural transitions: the putative cell entry states of poliovirus*. Journal of Virology, 2000. **74**(3): p. 1342-1354.
45. Speir, J.A., et al., *Enhanced Local Symmetry Interactions Globally Stabilize a Mutant Virus Capsid That Maintains Infectivity and Capsid Dynamics*. Journal of Virology, 2006. **80**(7): p. 3582-3981.
46. Tsang, S.K., et al., *Kinetic analysis of the effect of poliovirus receptor on viral uncoating: the receptor as a catalyst*. Journal of Virology, 2001. **75**(11): p. 4984-4989.
47. Canady, M.A., et al., *Large conformational changes in the maturation of a simple RNA virus, Nudaurelia capensis  $\omega$  virus (N $\omega$ V)*. Journal of Molecular Biology, 2000(299): p. 573-584.
48. Taylor, D.J., et al., *Correlation of chemical reactivity of Nudaurelia capensis  $\omega$  virus with a pH-induced conformational change*. Chem. Commun., 2003: p. 2770–2771.
49. Lee, K.K., et al., *Cooperative Reorganization of a 420 Subunit Virus Capsid*. Journal of Molecular Biology, 2005(352): p. 723-735.
50. Wikoff, W.R., et al., *Time-resolved molecular dynamics of bacteriophage HK97 capsid maturation interpreted by electron cryo-microscopy and X-ray crystallography*. Journal of Structural Biology, 2006: p. 300-306.
51. Lee, K.K., et al., *Evidence that a Local Refolding Event Triggers Maturation of HK97 Bacteriophage Capsid*. Journal of Molecular Biology, 2004. **340**(3): p. 419-433.
52. Freddolino, P.L., et al., *Molecular dynamics simulations of the complete satellite tobacco mosaic virus*. Structure, 2006. **14**(3): p. 437-449.
53. Speelman, B., B.R. Brooks and C.B. Post, *Molecular Dynamics Simulations of Human*

- Rhinovirus and an Antiviral Compound*. . Biophysical Journal, 2001. **80**: p. 121-129.
54. Durup, J., *Protein Molecular Dynamics Constrained to Slow Modes. Theoretical Approach Based on a Hierarchy of Local Modes with a Set of Holonomic Constraints: The Method and Its Tests on Citrate Synthase*. J Phys Chem, 1991. **95**: p. 1817-1829.
  55. Tuckerman, M.E. and B.J. Berne, *Molecular-Dynamics in Systems with Multiple Time Scales - Systems with Stiff and Soft Degrees of Freedom and with Short and Long-Range Forces*. Journal of Chemical Physics, 1991. **95**(11): p. 8362-8364.
  56. Tuckerman, M.E., B.J. Berne, and G.J. Martyna, *Molecular-Dynamics Algorithm for Multiple Time Scales - Systems with Long-Range Forces*. Journal of Chemical Physics, 1991. **94**(10): p. 6811-6815.
  57. Askar, A., B. Space, and H. Rabitz, *Subspace Method for Long-Time Scale Molecular-Dynamics*. Journal of Physical Chemistry, 1995. **99**(19): p. 7330-7338.
  58. Reich, S., *Smoothed dynamics of highly oscillatory Hamiltonian systems*. Physica D, 1995. **89**: p. 28-42.
  59. Phelps, D.K. and C.B. Post, *A Novel Basis for Capsid Stabilization by Antiviral Compounds*. Journal of Molecular Biology, 1995. **254**(4): p. 544-551.
  60. Phelps, D.K., P.J. Rossky, and C.B. Post, *Influence of an antiviral compound on the temperature dependence of viral protein flexibility and packing: a molecular dynamics study*. Journal of Molecular Biology, 1998. **276**(2): p. 331-337.
  61. Elezgaray, J. and Y.H. Sanejouand, *Modeling large-scale dynamics of proteins*. Biopolymers, 1998. **46**(7): p. 493-501.
  62. Elezgaray, J. and Y.H. Sanejouand, *Modal dynamics of proteins in water*. Journal of Computational Chemistry, 2000. **21**(14): p. 1274-1282.
  63. Feenstra, K.A., B. Hess, and H.J.C. Berendsen, *Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems*. Journal of Computational Chemistry, 1999. **20**(8): p. 786-798.
  64. Sorensen, M.R. and A.F. Voter, *Temperature-accelerated dynamics for simulation of infrequent events*. J Chem. Phys, 2000. **112**(21): p. 9599-9606.
  65. Chun, H.M., et al., *MBO(N)D: A multibody method for long-time molecular dynamics simulations*. Journal of Computational Chemistry, 2000. **21**(3): p. 159-184.
  66. Isea, R., C. Aponte, and R. Cipriani, *Can the RNA of the cowpea chlorotic mottle virus be released through a channel by means of free diffusion? A test in silico*. Biophysical Chemistry, 2004. **107**(2): p. 101-106.
  67. Harries, D., S. May, W.M. Gelbart and A. Ben-Shaul, *Structure, Stability and Thermodynamics of Lanellar DNA-Lipid Complexes*. Biophysical Journal, 1998. **75**: p. 159.
  68. Arkhipov, A., P.L. Freddolino, and K. Schulten, *Stability and dynamics of virus capsids described by coarse-grained modeling*. Structure, 2006(14): p. 1767-1777.
  69. Tama, F. and C.L. Brooks, *The mechanism and pathway of pH induced swelling in cowpea chlorotic mottle virus*. Journal of Molecular Biology, 2002. **318**(3): p. 733-747.
  70. Tama, F. and C.L. Brooks, *Diversity and identity of mechanical properties of*

- icosahedral viral capsids studied with elastic network normal mode analysis*. Journal of Molecular Biology, 2005. **345**(2): p. 299-314.
71. van Vlijmen, H.W.T. and M. Karplus, *Normal mode analysis of large systems with icosahedral symmetry: Application to (Dialanine)(60) in full and reduced basis set implementations*. Journal of Chemical Physics, 2001. **115**(2): p. 691-698.
  72. van Vlijmen, H.W.T. and M. Karplus, *Normal mode calculations of icosahedral viruses with full dihedral flexibility by use of molecular symmetry*. Journal of Molecular Biology, 2005. **350**(3): p. 528-542.
  73. Zhang, D.Q., et al., *Electrostatic interaction between RNA and protein capsid in cowpea chlorotic mottle virus simulated by a coarse-grain RNA model and a Monte Carlo approach*. Biopolymers, 2004. **75**(4): p. 325-337.
  74. Konecny, R., et al., *Electrostatic properties of cowpea chlorotic mottle virus and cucumber mosaic virus capsids*. Biopolymers, 2006. **82**(2): p. 106-120.
  75. Phillips, J.C., et al., *Scalable molecular dynamics with NAMD*. Journal of Computational Chemistry, 2005(26): p. 1781-1802.
  76. Liu, P., S. Izvekov, and G.A. Voth, *Multiscale coarse-graining of monosaccharides*. Journal of Physical Chemistry B, 2007. **111**(39): p. 11566-11575.
  77. Chu, J.W., et al., *Emerging methods for multiscale simulation of biomolecular systems*. Molecular Physics, 2007. **105**(2-3): p. 167-175.
  78. Izvekov, S. and G.A. Voth, *A multiscale coarse-graining method for biomolecular systems*. Journal of Physical Chemistry B, 2005. **109**(7): p. 2469-2473.
  79. Miao, Y. and P.J. Ortoleva, *Collective Modes and Order Parameters for Multiscale Simulation of Many-Particle Systems*. Physical Review Letters, 2009. **Submitted**.
  80. Sayyed-Ahmad, A., K. Tuncay, and P.J. Ortoleva, *Efficient solution technique for solving the Poisson-Boltzmann equation*. J Comput Chem, 2004. **25**(8): p. 1068-74.
  81. Sayyed-Ahmad, A., Y. Miao, and P.J. Ortoleva, *Poisson-Boltzmann Theory for Bionanosystems*. Communications in Computational Physics, 2007. **3**(5): p. 1100-1116.
  82. Roy, A.S., Y. Miao, and P. Ortoleva, *Multiscale Poisson-Boltzmann Approach to Nanosystem Electrostatics*. Journal of Computational Chemistry, 2008. **In Review**.
  83. Zwanzig, R., *Nonequilibrium Statistical Mechanics*. Oxford University Press, New York, 2001.
  84. Whitfield, T.W. and J.E. Straub, *Gravitational smoothing as a global optimization strategy*. Journal of Computational Chemistry, 2002. **23**(11): p. 1100-1103.
  85. Czermanski, R. and R. Elber, *Self-Avoiding Walk between 2 Fixed-Points as a Tool to Calculate Reaction Paths in Large Molecular-Systems*. International Journal of Quantum Chemistry, 1990: p. 167-186.
  86. Elber, R. and M. Karplus, *A Method for Determining Reaction Paths in Large Molecules - Application to Myoglobin*. Chemical Physics Letters, 1987. **139**(5): p. 375-380.
  87. Pratt, L.R., *A Statistical-Method for Identifying Transition-States in High Dimensional Problems*. Journal of Chemical Physics, 1986. **85**(9): p. 5045-5048.
  88. Elber, R., et al., *Bridging the Gap between Long Time Trajectories and Reaction*

- Pathways*. Advances in Chemical Physics, Vol 126, 2003. **126**: p. 93-129.
89. Bolhuis, P.G., et al., *Transition path sampling: Throwing ropes over rough mountain passes, in the dark*. Annual Review of Physical Chemistry, 2002. **53**: p. 291-318.
  90. Crooks, G.E. and D. Chandler, *Efficient transition path sampling for nonequilibrium stochastic dynamics*. Physical Review E, 2001. **64**(2): p. -.
  91. Faradjian, A.K. and R. Elber, *Computing time scales from reaction coordinates by milestoning*. Journal of Chemical Physics, 2004. **120**(23): p. 10880-10889.
  92. Elber, R., A. Ghosh, and A. Cardenas, *Long time dynamics of complex systems*. Accounts of Chemical Research, 2002. **35**(6): p. 396-403.
  93. Elber, R., *Long-timescale simulation methods*. Current Opinion in Structural Biology, 2005. **15**(2): p. 151-156.
  94. Chandrasekhar, S., *Dynamical Friction. I. General Considerations: the Coefficient of Dynamical Friction*. Astrophysical J., 1943. **97**: p. 255-262.
  95. Deutch, J.M. and I. Oppenheim, *The Concept of Brownian Motion in Modern Statistical Mechanics*. 1987, Faraday Discuss. Chem. Soc: London. p. 1-20.
  96. Deutch, J.M., et al., *Light-Scattering from Systems with Chemical Oscillations and Dissipative Structures*. Journal of Chemical Physics, 1972. **57**(10): p. 4327-&.
  97. Shea, J.E. and I. Oppenheim, *Fokker-Planck equation and Langevin equation for one Brownian particle in a nonequilibrium bath*. Journal of Physical Chemistry, 1996. **100**(49): p. 19035-19042.
  98. Shea, J.E. and I. Oppenheim, *Fokker-Planck equation and non-linear hydrodynamic equations of a system of several Brownian particles in a non-equilibrium bath*. Physica A, 1997. **247**(1-4): p. 417-443.
  99. Ortoleva, P.J., *Nanoparticle dynamics: a multiscale analysis of the Liouville equation*. J Phys Chem B, 2005. **109**(45): p. 21258-66.
  100. Peters, M.H., *Fokker-Planck equation, molecular friction, and molecular dynamics for Brownian particle transport near external solid surfaces*. Journal of Statistical Physics, 1999. **94**(3-4): p. 557-586.
  101. Peters, M.H., *Fokker-Planck equation and the grand molecular friction tensor for coupled rotational and translational motions of structured Brownian particles near structured surfaces*. Journal of Chemical Physics, 1999. **110**(1): p. 528-538.
  102. Pankavich, S., et al., *Stochastic dynamics of bionanosystems: Multiscale analysis and specialized ensembles*. J Chem Phys, 2008. **128**(23): p. 234908.
  103. Shreif, Z. and P. Ortoleva., *Curvilinear All-Atom Multiscale (CAM) Theory of Macromolecular Dynamics*. J. Stat. Phys., 2008(130): p. 669-685.
  104. Chang, R., G.S. Ayton, and G.A. Voth, *Multiscale coupling of mesoscopic- and atomistic-level lipid bilayer simulations*. Journal of Chemical Physics, 2005. **122**(24).
  105. Ayton, G.S. and G.A. Voth, *Multiscale simulation of transmembrane proteins*. Journal of Structural Biology, 2007. **157**(3): p. 570-578.
  106. Ayton, G.S., W.G. Noid, and G.A. Voth, *Multiscale modeling of biomolecular systems: in serial and in parallel*. Current Opinion in Structural Biology, 2007. **17**(2): p. 192-198.

107. Kevrekidis, I.G., C.W. Gear, and G. Hummer, *Equation-free: The computer-aided analysis of complex multiscale systems*. Aiche Journal, 2004. **50**(7): p. 1346-1355.
108. Hummer, G. and I.G. Kevrekidis, *Coarse molecular dynamics of a peptide fragment: Free energy, kinetics, and long-time dynamics computations*. Journal of Chemical Physics, 2003. **118**(23): p. 10762-10773.
109. Lange, O.F. and H. Grubmüller, *Can principal components yield a dimension reduced description of protein dynamics on long time scales?* Journal of Physical Chemistry B, 2006. **110**(45): p. 22842-22852.
110. Balsera, M.A., et al., *Principal component analysis and long time protein dynamics*. Journal of Physical Chemistry, 1996. **100**(7): p. 2567-2572.
111. Kitao, A. and N. Go, *Investigating protein dynamics in collective coordinate space*. Current Opinion in Structural Biology, 1999. **9**(2): p. 164-169.
112. Clarage, J.B., et al., *A Sampling Problem in Molecular-Dynamics Simulations of Macromolecules*. Proceedings of the National Academy of Sciences of the United States of America, 1995. **92**(8): p. 3288-3292.
113. Zhang, Z.Y. and W. Wriggers, *Coarse-Graining Protein Structures With Local Multivariate Features from Molecular Dynamics*. Journal of Physical Chemistry B, 2008. **112**(44): p. 14026-14035.
114. Dey, B.K., H. Rabitz, and A. Askar, *Optimal reduced dimensional representation of classical molecular dynamics*. Journal of Chemical Physics, 2003. **119**(11): p. 5379-5387.
115. Steven Hayward, A.K.N.G. and omacr, *Harmonicity and anharmonicity in protein dynamics: A normal mode analysis and principal component analysis*. Proteins: Structure, Function, and Genetics, 1995. **23**(2): p. 177-186.
116. R. M. Levy, A.R.S.W.K.O.J.A.M., *Quasi-harmonic method for studying very low frequency modes in proteins*. Biopolymers, 1984. **23**(6): p. 1099-1112.
117. Bradley, M.J., P.T. Chivers, and N.A. Baker, *Molecular Dynamics Simulation of the Escherichia coli NikR Protein: Equilibrium Conformational Fluctuations Reveal Interdomain Allosteric Communication Pathways*. Journal of Molecular Biology, 2008. **378**(5): p. 1155-1173.
118. Jaqaman, K. and P.J. Ortoleva, *New space warping method for the simulation of large-scale macromolecular conformational changes*. J Comput Chem, 2002. **23**(4): p. 484-91.
119. Dunweg, B. and K. Kremer, *Molecular-Dynamics Simulation of a Polymer-Chain in Solution*. Journal of Chemical Physics, 1993. **99**(9): p. 6983-6997.
120. Kopf, A., B. Dunweg, and W. Paul, *Dynamics of polymer "isotope" mixtures: Molecular dynamics simulation and Rouse model analysis*. Journal of Chemical Physics, 1997. **107**(17): p. 6945-6955.
121. Liu, B. and B. Dunweg, *Translational diffusion of polymer chains with excluded volume and hydrodynamic interactions by Brownian dynamics simulation*. Journal of Chemical Physics, 2003. **118**(17): p. 8061-8072.

122. Uvarov, A. and S. Fritzsche, *Friction of N-bead macromolecules in solution: Effects of the bead-solvent interaction*. Phys. Rev. E, 2006. **73**: p. 011111.
123. Pankavich, S., Z. Shreif, and P.J. Ortoleva, *Multiscaling for Classical Nanosystems: Definition of the Smoluchowski and Fokker-Planck Equations*. Physica A, 2008(387): p. 4053-4069.
124. Peters, M.H., *The Smoluchowski diffusion equation for structured macromolecules near structured surfaces*. Journal of Chemical Physics, 2000. **112**(12): p. 5488-5498.
125. Uvarov, A. and S. Fritzsche, *Effects of the bead-solvent interaction on the dynamics of macromolecules, 1 - The dumbbell molecule*. Macromolecular Theory and Simulations, 2004. **13**(3): p. 241-256.
126. Jacrot, B., *Studies on the assembly of a spherical plant virus : II. The mechanism of protein aggregation and virus swelling*. Journal of Molecular Biology, 1975. **95**(3): p. 433-446.
127. Tsang, S.K., et al., *Stabilization of poliovirus by capsid-binding antiviral drugs is due to entropic effects*. Journal of Molecular Biology, 2000. **296**(2): p. 335-340.
128. Humphrey, W., A. Dalke, and K. Schulten, *VMD: Visual molecular dynamics*. Journal of Molecular Graphics, 1996. **14**(1): p. 33-&.
129. Mackerell, A.D., et al., *Self-Consistent Parameterization of Biomolecules for Molecular Modeling and Condensed Phase Simulations*. Faseb Journal, 1992. **6**(1): p. A143-A143.
130. MacKerell, A.D., et al., *All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins*. J. Phys. Chem. B, 1998. **102**(18): p. 3586-3616.
131. Speir, J.A., et al., *Structures of the native and swollen forms of cowpea chlorotic mottle virus determined by X-ray crystallography and cryo-electron microscopy*. Structure, 1995. **3**(1): p. 63-78.
132. Ortoleva, P., *Nonlinear Chemical Waves*. 1992, NY: Wiley.
133. Ortoleva, P., Y. Chen and W. Chen, *Agates, geodes, concretions and orbicules: Self-organized zoning and morphology*, in *Fractals and dynamic systems in geoscience*, Jörn H. Kruhl, Editor. 1994, Springer-Verlag: NY. p. 283-305.
134. Kadau, K., et al., *Microscopic view of structural phase transitions induced by shock waves*. Science, 2002. **296**(5573): p. 1681-4.
135. Catti, M. and R.M. Ibberson, *Order-disorder of the hydronium ion and low-temperature phase transition of (H3O)Zr-2(PO4)(3) NASICON by neutron diffraction*. Journal of Physical Chemistry B, 2002. **106**(46): p. 11916-11921.
136. Guerin, T. and R. Bruinsma, *Theory of conformational transitions of viral shells*. Physical Review E, 2007. **76**.
137. Caspar, D.L.D., *Movement and Self-Control in Protein Assemblies - Quasi-Equivalence Revisited*. Biophysical Journal, 1980. **32**(1): p. 103-138.
138. Conway, J.F., et al., *Virus maturation involving large subunit rotations and local refolding*. Science, 2001. **292**(5517): p. 744-8.
139. Rader, A.J., D.H. Vlad, and I. Bahar, *Maturation dynamics of bacteriophage HK97 capsid*. Structure, 2005. **13**(3): p. 413-421.



140. Klug, W.S., et al., *Failure of viral shells*. Physical Review Letters, 2006. **97**(22): p. -.
141. Jorgensen, W.L., et al., *Comparison of Simple Potential Functions for Simulating Liquid Water*. Journal of Chemical Physics, 1983. **79**(2): p. 926-935.
142. Taylor, D.J., et al., *Correlation of chemical reactivity of Nudaurelia capensis omega virus with a pH-induced conformational change*. Chem Commun (Camb), 2003(22): p. 2770-1.
143. Henzler-Wildman, K. and D. Kern, *Dynamic personalities of proteins*. Nature, 2007. **450**(7172): p. 964-972.
144. Lindahl, E. and O. Edholm, *Spatial and energetic-entropic decomposition of surface tension in lipid bilayers from molecular dynamics simulations*. Journal of Chemical Physics, 2000. **113**(9): p. 3882-3893.
145. Lavelle, L., J.P. Michel, and M. Gingery, *The disassembly, reassembly and stability of CCMV protein capsids*. Journal of Virological Methods, 2007. **146**(1-2): p. 311-316.
146. Beardsley, R.L., Running, W. E., Reilly, J. P., *Probing the structure of Caulobacter crescentus ribosome with chemical labeling and mass spectrometry*. J. Proteome Res., 2006. **5**: p. 2935-2946.
147. Basavappa, R., et al., *Role and Mechanism of the Maturation Cleavage of Vpo in Poliovirus Assembly - Structure of the Empty Capsid Assembly Intermediate at 2.9-Angstrom Resolution*. Protein Science, 1994. **3**(10): p. 1651-1669.
148. Shreif, Z., et al., *Enveloped viruses understood via multiscale simulation: Computer-aided vaccine design*. Springer Netherlands, 2008. **15**(363-380).

# Yinglong Miao

Center for Cell and Virus Theory

Department of Chemistry, Indiana University

Address: 800 E Kirkwood Ave Room C203A, Bloomington, IN 47405

Phone: 812-856-0981(o) 812-272-8196(c) E-mail: yimiao@indiana.edu

## EDUCATION

---

### Indiana University

Bloomington, IN

#### *PhD in Theoretical Chemistry*

08/09

- Thesis: "All-atom Multiscale Computational Modeling of Viral Dynamics"
- Courses: statistical mechanics, quantum mechanics, virology, numerical analysis, scientific computing (GPA: 3.79/4.00)

### University of Science and Technology of China

Hefei, China

#### *Bachelor of Science in Chemistry*

06/04

- Thesis: "Theoretical Study of Electron Transport through Single Molecules with TranSIESTA-C"
- Courses: inorganic, organic, analytical, physical and biological chemistry, chemical labs, calculus, linear algebra, physics, physical labs, English, Chinese philosophy and imported culture, public administration (GPA: 3.64/4.00)

#### *Bachelor of Engineering in Computer Science and Technology (Dual Degree)*

06/04

- Thesis: "A C++ program for data collection and analysis proving Cooley-Tukey Fast Fourier Transform algorithm"
- Courses: database system, computer network technology and applications, operating systems, object-oriented system design and analysis and C++, digital signal processing (GPA: 3.68/4.00).

## AWARDS AND HONORS

- 
- "Oak Ridge Institute for Science and Education (ORISE) Participant Fellowship", 2007-2008
  - "Guanghua Education Scholarship", 2002-2003, USTC, China
  - "Outstanding Student Scholarship" for years 2002, 2001, 2000, USTC, China

## RESEARCH INTERESTS

- 
- Theoretical and computational methods for simulating biomolecular systems
  - Computer-aided drug design and structure-based molecular modeling
  - Multiscale theory and modeling of biological nanosystems
  - Protein-ligand binding, molecular docking and protein folding
  - High-performance parallel computing

## RESEARCH EXPERIENCE

---

### Indiana University

Bloomington, IN

#### *Research Assistant at Center for Cell and Virus Theory*

08/04 – 08/09

Advisor: Prof. Peter J. Ortoleva

- Coded a Poisson-Boltzmann solver to investigate viral electrostatic properties and taught it to a new graduate student
- Developed all-atom multiscale analysis (AMA) theory for nanosystem dynamics, constructed a novel molecular dynamics/order parameter extrapolation (MD/OPX) computational model and simulated viruses and other bionanosystems to predict and guide experiments through collaboration with virologists and biochemists
- Explained MD/OPX code and simulation results to two newly-joined Post-docs and helped with project starting
- Initiated MD/OPX application project on “allosteric switch in *Staphylococcus aureus* CzrA Zn sensor proteins”
- Instructed and coordinated with group website developer to construct a web portal for MD/OPX
- Delivered MD/OPX as open-source code to the chemical and biophysical research community via SimTk

#### *Oak Ridge Institute for Science and Education Participant*

07/07 - 06/08

- Developed nanosystem simulator (NanoX) based on MD/OPX for simulating large nanosystems with US Air Force Research Laboratory and introduced NanoX to the research community by installing it on Air Force supercomputer

### University of Science and Technology of China

Hefei, China

#### *Thesis of BS in Chemistry*

08/03 - 06/04

Advisor: Dr. Qunxiang Li

- Simulated the electron transport through two isoelectron molecules, 1,2-Bis(4'-mercaptophenyl) diazene and 1,2-Bis(4'-mercaptophenyl) ethene, that are sandwiched between two Au(111) surfaces with TranSIESTA-C, a first principles electronic structure program based on density functional theory (DFT) combined with non-equilibrium Green's function technique,
- Compared the electron transport properties of the two molecules
- Revealed their negative differential resistance (NDR) phenomenon at high voltage.

#### *Thesis of BE in Computer Science and Technology*

09/03 - 01/04

Advisor: Prof. Jiecheng Lu

- Designed and wrote a C++ program to gather data with a PS-2116 A/D card, processed the data using Fast Fourier Transform (FFT), and drew both the time signal graph of the data and its frequency spectrum that proves the radix-2 decimation-in-time Cooley-Tukey FFT algorithm.

#### *University Students' Research Project Participant*

09/02 - 09/03

Advisor: Prof. Wenming Chen

- Synthesized specialized macromolecular superabsorbents that are capable of absorbing distilled water of ~1000 times its original weight and investigated various properties, especially the anti-photodegradation under ultraviolet radiation.

## TEACHING EXPERIENCE

---

Indiana University

Bloomington, IN

*Associate Instructor for the Department of Chemistry*

Discussion on "Introduction to Physical Chemistry" 09/08 - 12/08

- Led discussions, graded homework, organized quizzes and held office hours to answer students' questions and enlightened students to grasp physical chemistry principles

Lab on "Introduction to Chemical Principles" 01/07 - 05/07

Lab on "Principles of Chemistry and Biochemistry I" 08/06 - 12/06

Lab on "Principles of Chemistry and Biochemistry II" 01/05 - 05/05

- Demonstrated experiments, explained reaction mechanisms, troubleshoot experimental problems, graded lab reports, proctored/graded exams and held office hours to address students' problems and help understanding the experiments

## JOURNAL PUBLICATIONS

---

1. **Miao, Y.** and P. J. Ortoleva (2006). All-Atom Multiscaling and New Ensembles for Dynamical Nanoparticles. *J. Chem. Phys.*, 125, 044901.
2. **Miao, Y.** and P. J. Ortoleva (2006). All-atom Multiscale Theory of Viral Structural Phase Transitions. *J. Chem. Phys.*, **125**, 214901.
3. Sayyed-Ahmad, A., **Y. Miao**, and P. Ortoleva (2007), Poisson-Boltzmann Theory for Bionanosystems. *Communications in Computational Physics*, **3**(5): 1100.
4. Huang, J., Q. Li, X. Wu, **Y. Miao**, J. Yang (2007), Transport Property of Two Isoelectronic Molecules. *International Journal of Nanoscience*, **5**(6): 841-846.
5. **Miao, Y.** and P. J. Ortoleva (2008), Molecular Dynamics/Order Parameter eXtrapolation for Bionanosystem Simulations. *J. Comp. Chem.*, **30**(3): 423-437.
6. Pankavich, S., **Y. Miao**, J. Ortoleva, Z. Shreif, P. Ortoleva (2008), Stochastic Dynamics of Bionanosystems: Multiscale Analysis and Specialized Ensembles. *J. Chem. Phys.*, **128**, 234908.
7. Pankavich, S., Z. Shreif, **Y. Miao**, P. J. Ortoleva (2009). Self-Assembly of Nanocomponents into Composite Structures: Derivation and Simulation of a Langevin Model. *J. Chem. Phys.*, 130(19): 194115.
8. Singharoy, A., **Y. Miao** and P. J. Ortoleva, Multiscale Poisson-Boltzmann Theory for Nanosystem Electrostatics. 2009, *J. Comp. Chem.*, in review.
9. **Miao, Y.** and P. J. Ortoleva, Collective Modes and Order Parameters for Multiscale Simulation of Many-Particle Systems. 2009, *Physical Review Letter*, submitted.
10. **Miao, Y.** and P. J. Ortoleva, Viral Structural Transition Mechanisms Revealed via Multiscale Molecular Dynamics/Order Parameter eXtrapolation Simulation. 2009, *Biopolymers*, in press.
11. **Miao, Y.**, J. E. Johnson and P. J. Ortoleva, All-atom Multiscale Simulation of CCMV Capsid Swelling, 2009, *J. Phys. Chem. B.*, submitted.

12. **Miao, Y.**, A. Singharoy, D. P. Giedroc, and P. J. Ortoleva, Allosteric Switch in Staphylococcus aureus CsrA Zinc Sensor proteins Simulated with MD/OPX. 2009, in preparation.

## CONFERENCE PRESENTATIONS

- 
- **Miao, Y.** and P. J. Ortoleva (2008). Poster: All-Atom Multiscale Computational Modeling of Viral Dynamics. *39<sup>th</sup> American Conference on Theoretical Chemistry*, Northwestern University, Evanston, IL
  - **Miao, Y.**, Z. Shreif and P. J. Ortoleva (2007). Talk: All-Atom Multiscale Computational Modeling of Viral Dynamics. *39<sup>th</sup> Midwest Theoretical Chemistry Conference*, Indiana University, Bloomington, IN

## ACTIVITIES

---

<b>Indiana University</b>	Bloomington, IN
<b><i>Weekly Hiking Organizer and Leader</i></b>	06/08 – 07/09
<ul style="list-style-type: none"><li>• Organize weekly hiking for a group of over 80 registered members from Indiana University and lead participants to explore hiking trails in Indiana state forests and parks</li></ul>	
<b><i>Volunteer Income Tax Assistance</i></b>	04/05
<ul style="list-style-type: none"><li>• Recognized by Department of the Treasury, Internal Revenue Service, Atlanta, GA</li><li>• Explained tax policies to international students and scholars, and helped the Internal Revenue Service to assist taxpayers with filing federal and Indiana state tax forms.</li></ul>	

## SKILLS

- 
- Operating systems: UNIX, LINUX, Windows
  - Programming: FORTRAN (extensive), C, C++; MPI, OpenMP; TCL scripting (extensive)
  - Software: Extensive usage of NAMD and VMD; Tinker, Gaussian, ChemOffice, LabView, RasMol; Maple, Mathematica, MATLAB; PHP&MySQL, HTML, Dreamweaver, Photoshop, Flash; Microsoft Office, Kaleigraph, Origin, Xmgrace, Gimp, AutoCAD, SQL
  - Languages: Chinese (native) and English (fluent).