Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

HHS Public Access
Author manuscript
J Phys Chem B. Author manuscript; available in PMC 2016 July 08.

Published in final edited form as:
J Phys Chem B. 2012 July 26; 116(29): 8423–8434. doi:10.1021/jp2119247.

Space Warping Order Parameters and Symmetry: Application to Multiscale Simulation of Macromolecular Assemblies

Abhishek Singharoy, **Harshad Joshi**, **Yinglong Miao**[*,†], and **Peter J. Ortoleva**
Center for Cell and Virus Theory, Department of Chemistry, Indiana University, Bloomington, Indiana 47405, United States

Abstract

Coarse-grained features of macromolecular assemblies are understood via a set of order parameters (OPs) constructed in terms of their all-atom configuration. OPs are shown to be slowly changing in time and capture the large-scale spatial features of macromolecular assemblies. The relationship of these variables to the classic notion of OPs based on symmetry breaking phase transitions is discussed. OPs based on space warping transformations are analyzed in detail as they naturally provide a connection between overall structure of an assembly and all-atom configuration. These OPs serve as the basis of a multiscale analysis that yields Langevin equations for OP dynamics. In this context, the characteristics of OPs and PCA modes are compared. The OPs enable efficient all-atom multiscale simulations of the dynamics of macromolecular assemblies in response to changes in microenvironmental conditions, as demonstrated on the structural transitions of cowpea chlorotic mottle virus capsid (CCMV) and RNA of the satellite tobacco mosaic virus (STMV).

Graphical abstract

---

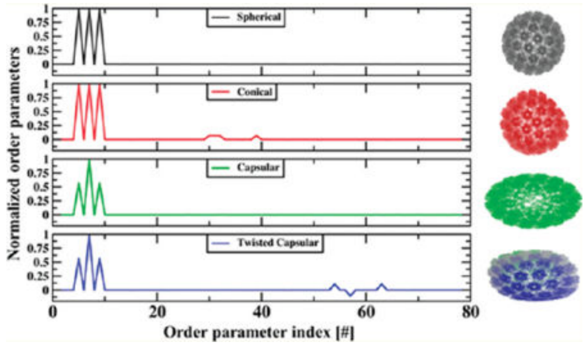
[*] **Corresponding Author:** yimiao@indiana.edu.
[†] **Present Address**
UT/ORNL Center for Molecular Biophysics, Department of Biochemistry & Cellular and Molecular Biology, University of Tennessee-Knoxville, Building 6011, Oak Ridge National Laboratory, Oak Ridge, TN 37831.

Special Issue: B: Macromolecular Systems Understood through Multiscale and Enhanced Sampling Techniques

Supporting Information
Comparison of DMS and NAMD simulation results with tables listing the simulation parameters, temperature dependence of OP evolution, and the dynamics of STMV RNA at high temperatures. This material is available free of charge via the Internet at http://pubs.acs.org.

Notes
The authors declare no competing financial interest.

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript



I. INTRODUCTION

A hallmark of macromolecular assemblies is the emergence of collective modes from rapidly fluctuating atomistic degrees of freedom (DoF). These systems exhibit dual macroscopic/microscopic behavior thereby reflecting the interplay of equilibrium and nonequilibrium processes across multiple time and length scales. Biologically relevant examples that display such coupling include processes affecting the structure and dynamics of macromolecular assemblies like viruses, ribosomes, liposomes, and intracellular granules.[1–4] These systems are typically composed of multimillion atoms. They function on length scales of nanometers involving processes that occur on time scales ranging from nanoseconds to milliseconds. While molecular dynamics (MD) has been widely used to simulate macromolecular structures at an atomistic level, the simulation time for large nanoscale assemblies has been limited to tens or sometimes few hundreds of nanoseconds.[5,6] Feasibility of such simulations also depends on the extent of computing resources available. Recently, billion atom MD simulations have been accomplished.[7,8] However, these simulations neglect Coulomb interactions, bonded forces, and the rapidly fluctuating hydrogen atoms. All the latter are central to biomolecular structure and dynamics. Thus, capturing bionanosystem behavior across diverse temporal and spatial scales presents great challenges in structural biology, fundamental mathematics and physics, and theoretical and computational chemistry.

Significant effort has been devoted to reduce the dimensionality of many-atom systems and accelerate their simulations by projecting the equation of motion in a low dimensional space.[9–13] This is accomplished via modeling a system in terms of its collective DoF. The number of such DoF is often much lesser than the all-atom ones. Thus, macromolecular models based only on collective modes involve tracking much smaller number of dynamical variables than the all-atom description. Consequently, computational cost of implementing these reduced dimensional models is moderate. A list of relevant approaches would include bead- and shape-based coarse-graining models,[14–16] rigid region decomposition,[17] symmetry constrained[18] and curvilinear coordinate[19] models, as well as principal component analysis (PCA)[20,21] and normal-mode analysis guided approaches.[22,23] These models have been successful in investigating structural transitions in a very rich set of biomolecular systems including BPTI, lysozyme, ligand-binding proteins,[24] trans-membrane proteins,[25] RNA segments,[20,26] GRoEL,[27] and viral protein capsids of different symmetries.[15,28] However, they suffer from one or more of the following difficulties in the

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

context of bionanosystem simulations: (1) characteristic variables are not slowly varying in time, (2) nonlinear motions like macromolecular twist is not readily accounted for, (3) internal dynamics, and hence inelasticity of collisions is neglected, (4) symmetry-breaking processes cannot be accounted for, (5) the forces involved must be calibrated for most new applications, and (6) generating intermediate all-atom trajectories for “on-the-fly” dimensionality reduction becomes very expensive for large systems.

A coarse-grained theory of macromolecular assemblies is statistical in character since specifying the coarse-grained variables leaves great uncertainty in the detailed all-atom state. Thus, the theory should provide an algorithm for evolving the coarse-grained variables and another for coevolving the probability of the detailed all-atom states. Acknowledging this multiscale perspective, we have discovered novel theoretical techniques that probe the cross-talk across scales in time and space, yet preserve the key all-atom aspects of the assembly dynamics.[29–35] The result is a set of stochastic equations for the evolution of coarse-grained variables, and those for constructing the coevolving probability of the all-atom states.

This is achieved via the introduction of a set of order parameters (OPs) that describe the overall organization of a system. These OPs capture changes in symmetry that follow large-scale structural transitions. Such transitions emerge from the interplay of long-range organization and order-destroying effects of thermal fluctuations. To account for this cross-talk between variables on different time and length scales, OPs enable generation of an ensemble of all-atom states. This ensemble, in turn, affects evolution of the overall structure through the thermal forces and diffusivities. Thus, emergence of new structures resulting from changes in conditions imposed on the macromolecular assemblies is probed. All these properties are critical for the practical implementation of a multiscale molecular dynamics/order parameter extrapolation (MD/OPX) approach[32,36,37] and more recently a fully self-consistent multiscale approach and software deductive multi-scale simulator (DMS).[38] These approaches have captured polyalanine folding from a linear to a globular state,[39] Ostwald’s ripening in nanocomposites,[31] pathways of structural transition and disassembly in virus capsids,[38] counterion induced collapse in viral RNA, and stability of RNA–protein complexes.[33]

In this article we review several examples of OPs in the context of modeling macromolecular assemblies (section II.A). A special class of macromolecular OPs, notably the space warping ones, is shown to account for slow collective DoF relegating the high frequency fluctuations to all-atom ensembles (section II.B). An attempt is made to place these coarse-grained variables within the classic notion of OPs that indicate symmetry breaking in phase transition theory[40] (section II.C). This analysis further reveals the physical significance of several of the space warping OPs (section II.D). These OPs serve as the basis of a multiscale simulation algorithm that captures the slow dynamics of macromolecular assemblies simultaneously preserving all-atom details (section II.E). In this context, the applicability of these OPs is compared to that of the PCA modes (section II.F). The OPs are used to simulate structural transitions in cowpea chlorotic mottle virus (CCMV) capsid and the free and protein bound states of RNA in satellite tobacco mosaic virus (STMV) over a range of temperature and salinity (section III). These simulations yield key insights on

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

macromolecular structural transitions and identify a regime of physical conditions over which the OP mediated multiscale simulations are applicable.

II. METHODOLOGY

A. Types of Order Parameters

OPs are coarse-grained variables characterizing the large-scale spatial organization of a system. Several types of OPs have been identified. Examples and the phenomena they have been used to describe are as follows.

- Scaled coordinates: Collective and single-particle behaviors in quantum systems,[41] and scaled center-of-mass coordinates for proteins in macromolecular assemblies.[42,43]
- Curvilinear coordinates: Macromolecular conformational dynamics.[19]
- Density-like variables: Release of drug molecules from a nanocapsule, the dynamics of enveloped viruses,[35] and liquid crystal phase transitions.[44,45]
- Space warping parameters: Overall size, shape and state of deformation of viruses and other macromolecular assemblies.[30,33,34]
- Subsystem OPs: The asymmetric motions of different subunits of a complex macromolecular assembly.[31,35]
- Hierarchical OPs: Disassembly/collapse dynamics of the icosahedral or other structures of viruses.[46]

Other examples of OPs in the theory of macromolecular structures commonly used are system diameter, end-to-end distance, radius of gyration, solvent accessible surface area (SASA), and measure of similarity to a reference structure in molecular biophysics (e.g., root-mean-square deviation of atomic positions between simulation and reference structures). However, these do not form a complete set, facilitate the construction of all-atom configurations, or evolve much slower than the typical time scale of atomistic fluctuations. Therefore, unlike the space-warping variables, these OPs cannot underlie a multiscale methodology (section II.E). Furthermore, space-warping OPs subsume the slowly varying parameters from the aforementioned list.[33] A central property of the space warping OPs is that they evolve slowly. The origins of slowness include (a) inertia associated with the coherent dynamics of many atoms evolving simultaneously, (b) migration over long distances, (c) stochastic forces that tend to cancel, and (d) presence of high free energy barriers involving collective motion. In the following, discussion on structural space warping OPs is extended.

B. Construction of Space Warping Order Parameters

We construct space warping OPs to capture coherent motions of many-atom systems as follows. Consider a macromolecular assembly described via the positions of its N constituent atoms $\vec{r_i}$ labeled $i = 1,\ldots, N$. In our approach, $\vec{r_i}$ is related to a reference position

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

$\vec{r}^{\,0}_{i}$. Deformation of space taking $\vec{r}^{\,0}_{i}$ to $\vec{r}_i$ is continuous and is used to introduce OP $\vec{\phi}_{\underline{k}}$ via the transformation

$$\vec{r}_i = \sum_{\underline{k}} u_{\underline{k}i}\vec{\phi}_{\underline{k}} \tag{1}$$

where $\vec{\phi}_{\underline{k}}$ denotes the $\underline{k}$th OP and the factor $u_{\underline{k}i}$ is defined in terms of a basis function $u_{\underline{k}}\left(\vec{r}^{\,0}_{i}\right)$ for reference position $\vec{r}^{\,0}_{i}$ of atom *i*. Index $\underline{k}$ labeling the $\vec{\phi}$ is a set of three integers $\{k_1k_2k_3\}$ such that $u_{\underline{k}i} \equiv u_{k_1k_2k_3}\left(\vec{r}^{\,0}_{i}\right) = u_{k_1}(X^0_i)u_{k_2}(Y^0_i)u_{k_3}(Z^0_i)$ is a product of Legendre polynomials of orders k_1, k_2, k_3 for the *X*, *Y*, *Z* components of $\vec{r}^{\,0}_{i}$ respectively. As the $\vec{\phi}_{\underline{k}}$ change, space is deformed, and so do the macromolecules embedded in it. Since we seek a dimensionality reduction, the number of $\vec{\phi}_{\underline{k}}$ is much less than the number *N* of atoms. Thus, we take a finite truncation of the $\underline{k}$ sum in (1); this necessitates introduction of a residual (denoted $\vec{\sigma}_i$) to correct the coherent deformation generated by the $\vec{\phi}_{\underline{k}}$. With this

$$\vec{r}_i = \sum_{\underline{k}} u_{\underline{k}i}\vec{\phi}_{\underline{k}} + \vec{\sigma}_i \tag{2}$$

An explicit expression for the $\vec{\phi}_{\underline{k}}$ is obtained by minimizing the mass-weighted square residuals $(m_1\sigma_1^2+\ldots+m_N\sigma_N{}^2)$ with respect to the $\vec{\phi}_{\underline{k}}$.[32] This yields

$$\vec{\phi}_{\underline{k}} = \frac{\sum_{i=1}^{N} m_i u_{\underline{k}i}\,\vec{r}_i}{\mu_{\underline{k}}};\mu_{\underline{k}} = \sum_{i=1}^{N} m_i u_{\underline{k}i}{}^2 \tag{3}$$

With the above formulation, dimensionality reduction for many-atom systems from the *N*-atom configuration to *M* OPs is achieved through

$$\vec{r}_i, \{i=1,\cdots,N\} \overset{u_{\underline{k}}}{\rightleftharpoons} \vec{\phi}_{\underline{k}}, \{\underline{k}=\underline{k}_1,\cdots,\underline{k}_M\} \tag{4}$$

Next, we make use of the Liouville equation to elucidate the rate of OP dynamics. The Liouville operator is defined $\mathrm{L} = -\sum_{i=1}^{N}\left(\vec{p}_i/m_i\right)\ \left(\partial/\partial\vec{r}_i\right) + \vec{F}_i\cdot\left(\partial/\partial\vec{p}_i\right)$, where $\vec{p}_i$ and $\vec{F}_i$ are the momentum of and net force on atom *i*. Given eq 3, one may compute $(\mathrm{d}\vec{\phi}_{\underline{k}})/\mathrm{d}t$ as $-\mathrm{L}\vec{\phi}_{\underline{k}}$. This yields

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$\frac{\mathrm{d}\vec{\phi}_{\underline{k}}}{\mathrm{d}t}=\frac{\sum_{i=1}^{N}u_{\underline{k}i}\,\vec{p}_i}{\mu_{\underline{k}}} \qquad (5)$$

By definition, the basis functions $u_{\underline{k}i}$ vary smoothly across the system. Thus, near equilibrium, linear combinations of rapidly fluctuating atomic momenta tend to cancel, thereby reducing the rate of OP evolution. As a result, these OPs can be extrapolated over longer periods in time relative to individual atoms. For the various choices of $\underline{k}$, evolution of the corresponding $\overline{\phi}$ tracks collective (correlated) motions on different time scales. With this, eq 5 provides mathematical basis for the simulation approach implemented in MD/OPX.[32]

C. Role of Symmetry

The concept of OPs originated in the theory of phase transitions. In that context, they are variables that are zero in one phase (usually above the critical point), and nonzero in another.[47] More generally, they change discontinuously across a first-order transition and their derivatives change across a second-order case indicating a change in the physical state of the system. For example, a magnetic system above the Curie temperature is isotropic, but anisotropy emerges below this temperature as atomic-scale magnets tend, on the average, to have a preferred direction.[40] Net magnetization of the system serves as an OP capturing the emergent order of atomic-scale magnets that underlie this symmetry breaking transition. Similarly, the space-warping OP here provides a framework that captures the emergent order observed in viruses and other macromolecular assemblies under appropriate conditions in the host medium. For example, if $u_{\underline{k}}\left(\vec{r}_i^{\,0}\right)$ is independent of i then the OP $\vec{\phi}_{\underline{k}}$ is proportional to the center-of-mass of the assembly.[46] Some of the space warping OPs defined in this way constitutes a strain tensor accounting for compression-extension-rotation, while others describe more complex deformations like tapering, twisting and bending (section II.D). Furthermore, the effective OP masses, $\mu_{\underline{k}}$ as defined in eq 3, decrease for larger values of $\underline{k}$. This suggests that OPs with higher $\underline{k}$ probe deformations of smaller regions in space. Thus, a model based on this set of OPs captures a spatially diverse range of coherent deformations through the various choices of $\underline{k}$. Thereby, such a model can describe the emergence of structural order and accompanying symmetry changes in macromolecular systems.

A typical example of macromolecular structural transition that resembles classical phase transition from isotropic to anisotropic states is now discussed. Consider a nanoscale assembly consisting of a spherical nanocore with a number of surface-attached viral capsid protein pentamers. In the absence of nanocore, pentamers often self-assemble into icosahedral structures of different T-numbers depending on the host medium conditions.[28] However, in the presence of the nanocore, arrangement of the surface-attached proteins may differ dramatically from those of the icosahedral structures in a nanocore free solution.[48] In particular, symmetry of the surface assemblies can be strongly influenced by the size and surface properties of the nanocore. If the core particle is of radius equals to the cavity of a T = 1 structure, then the pentamers would likely assemble in a T = 1 symmetry.[48] However, if

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

the core particle diameter is increased beyond the cavity size (but less than the diameter of the next T number structure) then a point is reached wherein bare spots appear on the core. Consequently, some preferred pentamer-to-neighbor-pentamer interactions are lost and these pentamers undergo Brownian motion over a range of temperatures. With this, the orientational order defining a $T = 1$ icosahedron can be disturbed and a symmetry breaking transition resulting in an assembly of randomly oriented pentamers is realized.

The above example can be placed more explicitly in the context of the classic relationship between symmetry breaking and emergence of OPs as follows. First, express the distribution of the pentamer density in terms of spherical harmonics and associated weights. At high temperature, the symmetric structure of pentamers is lost and hence the surface density is uniform (wherein weights of all harmonics except those for the lowest order ones are zero). Depending on the structure that emerges, at lower temperature, the weights of certain harmonics can depart from their values in the uniform state. In this case, weights of the spherical harmonics following the emergent geometry pattern (e.g., $T = 1$ icosahedron) serve as OPs in a manner similar to that used in the theory of ferromagnetism. When $u_{\underline{k}}\left(\vec{r}_i^{\,0}\right)$ are spherical harmonics, the $\vec{\phi}_{\underline{k}}$ in eq 3 provides a way to capture this transition. From this example it is seen, the space warping OPs are directly analogous to those appearing in the classical theory of phase transition for cases where simple symmetry can be identified.

The classical phase transition theories, like that for magnetization, are built on the properties of infinite systems, e.g., renormalization group concepts.[49] In contrast, macromolecular assemblies are finite, in fact small in extent and hence cannot completely follow the theory of macroscopic phase transitions. Furthermore, macromolecular assemblies can be in conformational states without a simple, readily identifiable symmetry, e.g., ribosomes. Nonetheless, as pH and other conditions in the host medium change, the system can switch to a different conformation.[50] Such a system experiences structural transition between two states, neither of which has a readily identifiable symmetry. This suggests that often OPs in macromolecules cannot be readily associated with the breaking of symmetry even if they signify a dramatic change of order. Therefore, other metrics are required to signal the emergence of new order in macromolecular systems when there are no readily identifiable symmetries involved.

For the space warping OP formulation of eq 3 such metrics include

- The onset of long-time tails in correlation functions of OP momenta indicating the coupling of a proposed set of OPs to emergent ones.[46]
- The systematic growth of residuals in eq 2 during structural transitions indicating the emergence of new types of organization from rapidly fluctuating atomistic processes.[46]

Numerical procedures for constructing the emergent OPs are provided in section III. With this, the space warping OPs can be used for the classical symmetry breaking type transitions in macromolecular systems and also provide an approach in cases where symmetries are not readily identifiable.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

D. Structural Characterization of Macromolecular Systems

Macromolecular assemblies display a rich array of structural organizations. For example, simple nonenveloped viruses often show an icosahedral arrangement of pentameric or hexameric protein subunits. However, under certain conditions they may exist in capsular, twisted and conical shapes.[28] Transition between these structures occurs via symmetry breaking pathways that account for changes in overall assembly organization. As shown below, the set of space warping OPs is rich enough that many types of symmetry are embedded in them. As in nature, observed symmetries for a given system emerge due to the underlying interatomic forces. This feature of the present approach enhances the capability to discover pathways of structural transition or self-assembly.[29,31]

Considering a $T = 7$ virus capsid as shown in Figure 1, it is demonstrated that specific combinations of space warping OPs capture the deformation of the symmetric $T = 7$ virus capsid into conical, capsular and twisted capsular forms. Transformations captured via these OPs are understood below in terms of a set of fundamental global and local deformations.[51] Consider the example of a tapering deformation (Figure 1). Take $U_{\underline{k}}$, $\underline{k} = 100, 110, 101$ to be x^0, x^0y^0, and x^0z^0, respectively. Neglecting residuals, eq 1 becomes $x_i=\phi_{100x}x_i^0+\phi_{100x}x_i^0y_i^0+\phi_{101x}x_i^0z_i^0$, and similarly for y_i and z_i (where x_i,y_i,z_i are the three Cartesian components of $r\rightharpoonup$ vector and $\varphi_{\underline{k}\alpha}$ is the α-th component of OP $\overrightarrow{\phi}_{\underline{k}}$). This relationship can be written in the tensorial form

$$\begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix} = J \begin{pmatrix} x_i^0 & \\ x_i^0 & y_i^0 \\ x_i^0 & z_i^0 \end{pmatrix} ; \; J = \begin{pmatrix} \phi_{100x} & \phi_{110x} & \phi_{101x} \\ \phi_{100y} & \phi_{110y} & \phi_{101y} \\ \phi_{100z} & \phi_{110z} & \phi_{101z} \end{pmatrix} \quad (6)$$

For the choice of $\varphi_{\underline{k}}$ as specified in Figure 1 (i.e., φ_{100x}, φ_{100y}, φ_{110x}, $\varphi_{110z} = 0$ and φ_{100x}, φ_{110y}, φ_{101z}, φ_{101y}, $\varphi_{101x} \neq 0$), J is the Jacobian matrix for tapering along the X axis.[51] Thus, these OPs capture structures that are tapered with respect to the reference configuration $\overrightarrow{r}_i^0$. Similar matrix transformations can be constructed using other combinations of OPs to explain the twisting and bending transitions of Figure 1. In this context, examples of OPs that enable motions like extension-compression-rotation were provided earlier.[39,46] Thus, the space warping variables are coarse-grained in character and, in addition, are rich enough to capture the emergence of overall order and symmetry breaking in macromolecular assemblies. With this, they enable a multiscale methodology as follows.

E. Deductive Multiscale Analysis

Equation 3 implies that for a given set of atomic positions the corresponding OPs $\overrightarrow{\phi}_{\underline{k}}$ are uniquely defined. However, the converse is not true, i.e., there exists multiple all-atom configurations consistent with a given set of OPs $\overrightarrow{\phi}_{\underline{k}}$. Thus, the OP construction scheme constitutes a many-to-one mapping from the all-atom to the coarse-grained description. As a consequence, the OP description retains overall structural information, losing all-atom

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

details. Atomistic structures are reconstructed via a procedure called deductive multiscale analysis (DMA) that evolves the OPs with an ensemble of all-atom configurations as follows.

The description adapted starts with the probability density ρ of the N atomic positions and momenta Γ. However, this formulation masks the underlying hierarchical organization of a macromolecular assembly. To address this, in DMA ρ is hypothesized to depend on Γ both directly, and via a set of OPs, indirectly. With this ansatz, a multiscale analysis of the Liouville equation yields sets of coupled Langevin equations for the OPs[31,52]

$$\frac{\mathrm{d}\vec{\phi}_{\underline{k}}}{\mathrm{d}\tau}=\sum_{\underline{k}'}\overset{\rightharpoonup}{\overset{\rightharpoonup}{D}}_{\underline{k}\underline{k}'}\vec{f}_{\underline{k}'}+\vec{\xi}_{\underline{k}} \qquad (7)$$

where $\tau = \varepsilon^2 t$ and ε is a smallness parameter, e.g., ratio of typical atomic mass to that of the entire assembly. The variance of noise $\vec{\xi}_{\underline{k}}$ is bound by diffusivity $\overset{\rightharpoonup}{\overset{\rightharpoonup}{D}}_{\underline{k}\underline{k}'}$. The factors $\overset{\rightharpoonup}{\overset{\rightharpoonup}{D}}_{\underline{k}\underline{k}'}$ are related to the ensemble average of correlation functions between hierarchical OP momenta $\vec{\Pi}_{\underline{k}}$ via

$$\overset{\rightharpoonup}{\overset{\rightharpoonup}{D}}_{\underline{k}\underline{k}'}=\frac{1}{\mu_{\underline{k}}\mu_{\underline{k}'}}\int_{-\infty}^{0}\mathrm{d}t_0\langle\vec{\Pi}_{\underline{k}}e^{-L_0t_0}\vec{\Pi}_{\underline{k}'}\rangle \qquad (8)$$

where $\mu_{\underline{k}}$ is the effective mass associated with $\varphi_{\underline{k}}$ as defined in eq 3 and $\mathrm{L}_0=-\sum_{i=1}^{N}(\vec{p}_i/m_i)(\partial/\partial\vec{r}_i)+\vec{F}_i\cdot(\partial/\partial\vec{p}_i)$ is the lowest order Liouville operator. The thermal-average force $\vec{f}_k$ is given by

$$\vec{f}_k=-\frac{\partial F}{\partial\vec{\phi}_{\underline{k}}}=\left\langle\sum_{i=1}^{N}u_{\underline{k}i}\,\vec{F}_i\right\rangle \qquad (9)$$

for OP-constrained Helmholtz free-energy F, where

$$F=-\frac{1}{\beta}\,ln\,\mathrm{Q}(\phi,\beta) \qquad (10)$$

$Q(\varphi,\beta)$ is the partition function constructed from configurations consistent with the set of $\vec{\phi}_{\underline{k}}$ (denoted φ collectively).

In the above formalism, thermal average forces $\vec{f}_k$ are constructed at each Langevin time step via Monte Carlo integration of atomic forces $\vec{F}_i$ obtained from interatomic force fields

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

(e.g., CHARMM[53]). Note that this does not require any assumption on the form of the dependence of the thermal average forces on the OPs; this is automatically incorporated because the all-atom ensembles used to carry out the thermal averaging are constructed for the values of the OPs at the given Langevin time step. Furthermore, this formalism accounts for the full impact of fluctuations $\overrightarrow{\xi}_k$ as the random forces in the Langevin equations are constructed to be consistent with the diffusivity factors $\overrightarrow{D}_{\underline{\mathrm{kk}}'}$.

The thermodynamic forces involved are small so that the state is captured by the slowly varying OPs. As the system evolves toward equilibrium the thermodynamic forces vanish and the system is fluctuation dominated. Both these cases are accounted for in our multiscale OP approach since the intensity of fluctuations is chosen to be consistent with the diffusion factors. Finally, if the thermodynamic forces are extremely large then the time scale of OP evolution approaches that of atomistic fluctuations. With this, the time scale separation between fast and slow variables is violated and the theory does not hold.

Inherent in our DMA approach is the capability to reconstruct atomistically resolved states given the evolving coarse-grained dynamics. Any coarse-grained theory carries an inherent uncertainty in the fine scale states.[54] DMA addresses this by providing the conditional probability density for the atomistic configurations given the instantaneous values of the OPs. These configurations are generated by a procedure denoted hybrid sampling in which the all-atom structures are reconstructed from the coarse-grained description via randomly varying the residuals $\overrightarrow{\sigma}_i$ at constant values of OPs (eq 2).[34] Therefore, the OPs constrain the ensemble of atomic states, while the latter determine the thermal average forces and diffusivity factors that control OP evolution (eq 7). Thus, the OPs imply a multiscale simulation algorithm that accounts for cross-talk between coarse-grained and atomic DoF. All these ideas are implemented as the DMS software.

F. Use of PCA and Legendre Basis Functions for Order Parameter Construction

Bionanosystems undergo large-scale conformational changes involving collective motions of strongly interacting clusters of atoms. Such structural changes often correspond to the functionally relevant motions of macromolecular systems.[4,21] Collective DoF can be observed through calculating the autocorrelations of the normalized 3*N*-dimensional atomic displacement vectors from consecutive time windows as

$$\Delta_{lm}=\Delta\overrightarrow{x}^{t_l}(\delta t)\cdot\Delta\overrightarrow{x}^{t_m}(\delta t) \quad (11)$$

where $\Delta\overrightarrow{x}^{t_l}(\delta t)$ denotes the normalized 3*N*-dimensional atomic displacement vector during the *l*-th time window of size δt for $l = 1,\ldots, L$.

There is a variety of coarse-grained variables in the literature that are used to analyze coherent motions in biomolecular systems (section I). Here, we test the suitability of using the basis vectors obtained from PCA for constructing the slow variables used in the multiscale analysis of section II.E. To address this issue, consider using the PCA modes (or eigenvectors) as the basis functions for constructing OPs (see the Appendix). Correlations

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

between the PCA derived basis functions U_k are compared to those of the Legendre polynomial ones u_{ki} at selected time intervals. Strong correlation in the long time behavior of these variables implies slow evolution of the functions used to construct the basis vector which, in turn, underlies coherent OP behavior. With this, define a correlation matrix **C** for the basis vectors via

$$C_{qk}=U_q^{t_m}\cdot U_k^{t_l} \quad (12)$$

where $U_k^{t_l}$ denotes the kth PCA or Legendre basis vector ($k = 1,\ldots, M$) from the l-th time window ($l = 1,\ldots, L$). In analogy with the PCA modes, the Legendre polynomials are constructed here using atomic displacements of a given time interval t_l (and not absolute positions as in eq 1, Appendix). The average structure required to calculate these PCA modes is changed with each time window, and similarly for the reference coordinates needed to construct OPs. This treatment of the Legendre polynomials enables a fair comparison with PCA modes in the context of analyzing basis function behavior as both variables are then expressed in terms of atomic displacements for a given time interval. This particular construction of u_{ki} is used for PCA comparisons only.

III. RESULTS AND DISCUSSION

A. Suitability of Coarse-Grained variables for Deductive Multiscaling

PCA has been applied to reduce the dimensionality of MD trajectories for analyzing large-scale structural changes.[21,55] Their behavior is compared with that of the space warping OPs for dimensionality reduction and deductive multiscaling. For demonstration, we choose the CCMV capsid as our model system because of its extensively studied structural transition phenomena.[56,57] The crystal structure of wild-type CCMV is solved at 3.2 Å resolution by X-ray crystallography.[57] Its capsid consists of 180 chemically identical protein subunits that form a 286-Å-diameter icosahedral shell with a $T = 3$ quasi-symmetry.

Macromolecular assemblies (e.g., virus capsids) evolve in stochastic fashion, and therefore simulating their long-time behavior should account for an ensemble of atomic positions and momenta. Thus, an analysis aimed at extracting low-frequency modes (e.g., space warping OPs or PCA) from MD should in principle be performed with an ensemble of trajectories. Here, 30 replica NAMD[58] runs with different random atomic velocity initializations are executed starting with the swollen CCMV capsid. Since the characteristic time of OP evolution is expected to be ~O(N) times longer than that of atomistic fluctuations (see eqs5 and 7), each MD trajectory is run for 10^5 fs (100 ps) for investigating the low-frequency modes.

While it is ideal to study the time evolution of all atoms in the system, it is impractical to perform PCA diagonalization of the large $3N \times 3N$ atomic positional covariance matrix for the 400 000 atom capsid. Thus, only data on 10% of C_α atoms (every tenth residue, 2940 in total) in the CCMV capsid backbone is extracted for our PCA study. We divide the 100 ps ensemble-averaged trajectory of swollen CCMV capsid into 10 time windows, each of

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

which is input to the MD analysis program carma[59] for PCA. Every calculation yields 8,820 eigenvectors and their eigenvalues. The first 20 PCA eigenfuctions are found to capture about 95% of structural changes in swollen CCMV capsid. We then choose these 20 modes from each time window and calculate their correlation matrix C from consecutive time windows using eq 12 to compare their similarities and thereby monitor coherence.

Figure 2a shows dot products of atomic displacement vectors $\Delta\overrightarrow{x}$, i.e., the Δ matrix expressed via eq 11 using 10 ps time windows. The overlap of $\Delta\overrightarrow{x}$ from neighboring time windows (the upper and lower diagonal entries) are found to be significant (about 0.6). As the distance between time windows increases, the $\Delta\overrightarrow{x}$ autocorrelations decrease gradually to about 0.2. These correlations imply that the capsid undergoes slowly evolving collective motions over time intervals of 10 ps or longer. Figure 2b shows the correlation matrix **C** of PCA basis vectors from the first two consecutive time windows calculated using eq 12. These PCA vectors are found to display small overlapping with their dot products distributed between −0.02 and +0.02. Basis vectors from other sets of consecutive time windows also give similar results. In contrast, Legendre polynomials of atomic displacements display significantly larger correlation between the basis vectors of the two consecutive time windows (similar values as the correlations of Δ) compared with that of the PCA (see Figure 2c). The antidiagonal entries have switching positive and negative values because of the odd and even properties of Legendre polynomials used.

Space warping basis vectors constructed with Legendre polynomials of atomic displacements at consecutive time intervals display large similarities implying coherence of the basis vectors u_{ki} as shown in Figure 2c, which, in turn, reflects the slowly evolving nature of the collective motions illustrated in Figure 2a. This suggest that the OPs constructed from these basis vectors are suitable to project system configuration over long time (10 ps or longer). Therefore, they can serve as a basis for the multiscale analysis introduced in section II.E. The comparison of Legendre with PCA derived basis vectors implies that the latter capture correlated motions on time scales lesser than 10 ps, not suitable to project system configuration over long time, and do not manifest the coherence in evolution required to construct OPs for deductive multiscaling and related simulations. Thus, temporal scales captured through these PCA modes are shorter than those described by OPs constructed from the Legendre polynomials of Figure 2. In the deductive multiscale formalism however, such local motions are captured through the construction of all-atom ensembles that probe events on time scales much smaller than that of OP evolution (section II.E). Alternatively, incorporation of more OPs in the coarse-grained description enables capturing short scale collective motions[32] (as discussed in sections III.B and III.C). Thus, the space warping OPs when embedded in our multiscale approach together with the implied quasi-equilibrium ensemble of OP constrained atomic configurations capture dynamics on shorter, similar, as well as longer scales than PCA modes. As a result, the OPs provide an ideal and versatile reduced description that enables the deductive multiscaling of N-atom macromolecular assemblies.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

B. Symmetry Breaking and CCMV Capsid Structural Transition

The slowly evolving nature of OPs suggests a multiscale MD/OPX approach to simulate bionanosystem dynamics.[32,36,37] In this approach, a set of replica MD runs with different velocity initializations is used to estimate the rate of OP change. Since OPs evolve coherently, they are readily extrapolated over long time periods using this rate through eq 5. Thus, the slow overall dynamics of the system is simulated. At every time step, the replica MD runs are repeated to reassess the rate of OP change that enables further extrapolation in time.

Biomolecular systems of nanoscale size provide ideal examples for illustrating the effectiveness of an OP based simulation methodology. MD/OPX was used to simulate the shrinkage of a swollen CCMV capsid in vacuum for 200 ns.[36] 3^3 OPs were constructed by using Legendre polynomials of atomic coordinates over order (0, 1, 2) in X, Y, and Z directions. Low-order Legendre polynomials were selected because they vary smoothly in space and thus can capture the overall nanoscale deformation of the virus capsid. Thirty 500 fs replica MDs were used for calculating the rate of OP evolution and extrapolating them to evolve the system. As the OPs are slowly varying in time, the length of MD runs used to compute the rate of their change are typically much smaller than the extrapolation time step. With this, adaptive timesteps in the range of 50–60 ps are achieved to extrapolate the OPs. These are 4 orders of magnitude larger than the typical MD timesteps (e.g., 1 fs). As shown in Figure 3a, the CCMV shrinkage 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 shrinkage. As a result, this viral structural transition starts locally and then propagates across the capsid, i.e., they proceed via intermediate states that are not constrained to the icosahedral symmetry of the initial and final states (i.e., $T = 3$).

Variables that capture the symmetry breaking nature of this CCMV capsid structural transition are the hexamer and pentamer orientation angles (Figure 3b). Their values change gradually during capsid shrinkage, indicating change in orientational order of the capsid subsystems. Insight into the nature of transitions involving these variables can be gained via analogy with the ferromagnetic phase transitions. For example, magnetic orientational order changes as temperature changes across the Curie value. Similarly, the capsid structural transition involves capsomere rotation as the pH and thus the protonation state of protein residues in the system changes.[37,56] However, unlike traditional examples in ferromagnetic transitions where the crystal structure remains intact, the orientational reordering of capsomeres accompanies shrinkage of the $T = 3$ capsid. This behavior, in turn, is similar to ferromagnetic phase transitions in alloys that simultaneously involve negative expansion of lattice parameters.[60] Thus, transitions involving change in capsomere orientations for a fixed T-number could occur as analogues to the change in lattice structure accompanying some ferromagnetic transitions.[61] Such structural transitions are also observed in molecular solids and liquid crystals[44] where eularian angles are used as orientational OPs.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

C. Structural Transition between States of Uncertain Symmetry

Traditionally, OPs are used to characterize macroscopic phase transitions wherein symmetry breaking is usually apparent since the symmetries of the states on either side of the transition are readily identified. While transitions in macromolecular systems often demonstrate large-scale structural reorganization, the symmetries involved can be less apparent (section II). Here, we use DMS[38] simulations to illustrate new OP related metrics that signal the change of order in macromolecular systems without readily identifiable symmetry. These simulations are based on the two-way transfer of structural information between OPs (coarse-grained description) and atomistic configurations (characterized by quasi-equilibrium probability densities) as described in section II.E. The demonstration system, RNA of STMV, is chosen because it (a) has interesting complexity, containing 949 nucleotides arranged in 30 double stranded helical stems joined via single-stranded loops, (b) is highly flexible, (c) expresses nonlinear motions and, (d) unlike CCMV, does not have a readily identifiable initial symmetry.[62] With this, applicability of the space warping OPs of section II as coarse-grained variables that efficiently probe highly complex motions is tested.

The initial state of the STMV RNA is taken to be when it resides inside the capsid with associated protein subunits. The simulated evolution follows after the capsid is removed instantaneously. In a 1:1 electrolyte like NaCl, the RNA initially expands, then shrinks due to electrostatic shielding by the diffusive counterion cloud, and finally fluctuates among atomistic states of similar energy.[33] All-atom configurations resulting from this simulation imply that the RNA tertiary structure is highly disrupted, although secondary structures remain after 50 ns.[33] In contrast, in a 2:1 electrolyte like $MgCl_2$, Mg^{2+} ions tightly bind to the RNA. Thus, secondary and tertiary structure is preserved during the 25 ns simulation (see Figure S1). This DMS predicted RNA stability in 0.3 M 2:1 electrolyte is in agreement with previous experimental and theoretical predictions[6,63] and reconfirmed here with NAMD simulation results at 310 K (Figure S2a).

To further test the robustness of DMS, the RNA simulation in 0.3 M $MgCl_2$ is repeated over a range of temperatures between 310 and 425 K. At each temperature 25 ns all-atom trajectories are obtained. Details of conditions and parameters used for these simulations are listed in Tables S1 and S2 in the Supporting Information. Increase in temperature results in the emergence of new collective motions. To capture these motions, OPs that were not included to the initial set of collective variables are added to the reduced description and subsequently evolved via Langevin equations (see section SI2 of the Supporting Information for details). Such OPs, denoted $\vec{\phi}_{\underline{k}}^{\text{new}}$ corresponds to values of $\underline{k}$ for which nonlinear local motions like tapering, twisting and bending are probed (Figure 4b). This illustrates that new space warping OPs can be readily added to capture complex motions as they emerge in response to change changes in the microenvironment.

Tracking the decay characteristic of OP velocity autocorrelation functions calculated by using eq 8 provides a criterion of completeness for a given set of OPs. In particular, appearance of long time tails in these correlation functions typically suggests there are additional OPs that couple to the set considered originally.[33,46] As these "missing" modes are accounted for via the introduction of additional OPs ($\vec{\phi}_{\underline{k}}^{\text{new}}$), the corresponding OP

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

velocity autocorrelation functions rapidly decay indicating completeness of the reduced description (Figure 4a). With this, the growth of residuals $\vec{\sigma}_i$ (Figure S3) and emergence of long tails in correlation functions (manifested during the construction of all-atom ensembles via hybrid sampling as described in section II.E) indicate the necessity of additional OPs that capture a shift in structural organization. Therefore, these quantities serve as OP related metrics that signal the change of order in macromolecular systems even if symmetry breaking is not readily identifiable.

The DMS results at 310 K and 425 K are benchmarked with those of 5 ns NAMD simulations. The OP trajectories obtained from these simulations are found to be in agreement (Figure S2); this validates the applicability of DMS simulations over a biologically relevant range of temperatures. However, as temperature increases the time scale of OP evolution decreases. This limits the size of Langevin OP timesteps and, subsequently, the applicability of the noninertial Langevin equations that underlie DMS simulations. Thus, simulations in the 310–425 K range are carried out using DMS; higher temperature simulations are performed using NAMD to capture the inertial regime that restricts DMS applicability (section SI2, in the Supporting Information). Comparison with MD at multiple temperatures ensures that DMS predicted results are free from artifacts. This also confirms that the predicted temperature behavior of the RNA is not affected by the switch in simulation methodology (DMS to MD) at 450 K.

The present investigation of RNA dynamics over a range of temperatures provides insights into a macromolecular phase transition. The essence of this transition is seen in Figure 5 where ensemble average values of three selected OPs are plotted versus temperature. The temperature-dependent behavior of average OP values suggests a phase transition type phenomenon occurs between 475 and 500 K. OPs undergo a drastic change in magnitude across this transition region (Figure 5). Observed increase in OPs implies expansion of the RNA structure. Thus, space warping OPs indicate a substantial shift in order between states of RNA in which symmetries are not distinct. Since these OPs capture overall RNA structure, analogous transition behavior is observed for other coarse-grained descriptors like the number of hydrogen bonds (Figure 5), radius of gyration (Figure S4), and end-to-end distance.

Most RNAs have a degree of heterogeneity in the nucleotide sequence. With this, one might expect heterogeneous nucleation sites for the transition. i.e., nucleotides are not all equally responsible for the structural changes in RNA. In the present simulations, the 14th and 19th nucleotide (ADE and URA, respectively) in each of the 30 RNA helices were found to be more thermo-labile, i.e., sensitive to changes in temperature. As shown in Figure 6a, enhanced motion of these nucleotides is reflected in significant shifts in their dihedral angles as the temperature changes. A Poisson–Boltzmann evaluation of electrostatic energies[64] shows that these two nucleotides are the most stable when RNA is embedded inside the virus capsid (Figure S5). Strong protein-nucleic acid interactions hold dihedral angles neighboring these nucleotides unchanged. However, when the capsid is removed, setting the RNA free as simulated here, the associated dihedrals change appreciably. Consequently, these dihedrals sample a range of values that evolve away from those of the protein-bound RNA (Figure 6a). As the transition zone is entered the motion of these thermo-labile centers increase until a

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

RNA-wide transition is expressed. For example, at these temperatures (475–500 K), the internucleotide hydrogen bond breaks providing additional DoF to the RNA helices (Figure 5). This, together with enhanced mobility of the thermo-labile centers, results in correlated nucleotide motions in their vicinity. Emergence of such motions like twisting or bending (Figure 4) disrupts secondary structure of the double stranded RNA stem and mediates the propagation of an instability front across the entire macromolecule (Figure S6) that transforms all the helices to coils (Figure 6c). Resulting shift in order of the overall RNA structure is indicated by the discontinuous temperature behavior of space warping OPs across the transition region (Figure 4). Thus, the thermo-labile nucleotides provide centers of nucleation that initiate the observed first-order like phase transition between the encapsidated and coiled states of STMV RNA. Root mean square (rms) fluctuations about the mean dihedrals for nucleotides of the RNA helices are presented (Figure 6b). Such fluctuations increase near the transition zone as expected if the analogy to macroscopic phase transitions is to hold.

In the above, the temperature-dependent behavior of STMV RNA is investigated from the space warping OPs, nucleation and fluctuation perspectives. Results suggest a first-order like phase transition occurs between 475 and 500 K. Such behavior reflects the interplay between local and overall structures of RNA. Although there is no experimental data available for the transition temperature of STMV RNA, earlier simulations on smaller RNA loops suggest a melting point between 420 to 430 K.[65] Given that the present system is much larger (i.e., 949 versus 14 nucleotides and an order or 2-fold greater number of dinucleotide bonds), the high transition temperature suggested here is not surprising. Besides, significant discrepancies between experimental and theoretical results on macromolecular structure and dynamics are well-known.[66] These differences can be attributed to factors such as the finite size of all-atom ensembles constructed at every Langevin time step for computing the thermal forces and diffusions (section II.E), possible incompatibility of force-fields with high temperature simulation,[65] discrepancies in thermal expansion coefficient of TIP3P water model,[67] and problems with simulating the counterion environment as realized in the experiments.[68] Nonetheless, RNA simulations in aqueous medium have been performed for temperatures in the range 300–700 K.[65,69] These simulations have provided insights into the thermorigidity and functionality of multiple RNA structures. Even though the above simulation inconsistencies might affect computing exact transition temperatures, the behavior of OPs across transition regions as reported here qualitatively follows previously observed trends.[66,69,70] With this, the space warping OPs used for macromolecular assemblies have properties analogous to those used in the theory of macroscopic phase transitions.

Our results also suggest that proteins on the inner surface of STMV capsid greatly enhance the stability of the thermo-labile nucleotides, i.e., suppress motion of the nucleation centers (see Figures 6a and S7). Consequently, viruses can survive in high temperatures (~400 K) without losing the RNA tertiary structure.[71] Thus, the free-energy minimizing structure of a virus can enable it to withstand severe temperatures, as commonly observed. At very high temperatures however, these protein–RNA interactions are lost and so the virus ultimately loses stability.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

D. Performance of Multiscale Techniques

Structural transitions in CCMV and RNA of STMV provide examples of phenomenon in which accounting for the cross-talk among multiple time and length scales becomes critical. Since both DMS and MD/OPX are based on the interplay of all-atom and coarse-grained variables, such multiscale methodologies naturally account for this cross-talk. Furthermore, application of these techniques leads to efficient computation of slow processes. For example, MD/OPX simulation with adaptive OP extrapolation timestepping is about 6 times faster than direct MD. Similarly, DMS is found to be about 11 times faster than NAMD at 310 K. This efficiency reduces by 4 fold at 425 K. This is expected as the OPs loose coherence with increase in temperature. However, a direct comparison with conventional MD run is not appropriate. This is because at each OP time step 200–300 all-atom structures consistent with the instantaneous OP values are obtained. Thus, DMS or MD/OPX simulation corresponds to an ensemble of 200–300 conventional MD runs, for which the timecourses of the spacing warping OPs are essentially the same, while the detailed atomistic configuration varies among members of the ensemble. Finally, a single MD run may not be representative of an ensemble of possible time courses, which, in contrast is automatically overcome in the all-atom multiscale approach.

IV. CONCLUSIONS

Space warping OPs capture key elements of symmetry breaking manifested in structural transitions of macromolecular assemblies. Their generality allows them to capture a variety of symmetries that emerge as a consequence of the underlying interatomic forces and conditions to which the system is subjected. The space warping OPs enable a multiscale analysis which accounts for the cross-talk between the coarse-grained and all-atom DoF, and characterizes an ensemble of atomic configurations coevolving with the OPs. In this context, spacing warping OPs together with the multiscale analysis capture a more diverse range of coherent motions than do PCA modes. These ideas are demonstrated via simulations of symmetry breaking accompanying CCMV capsid shrinkage, and a first order structural transition in RNA of STMV. The multiscale methodology presented here is robust to a range of salinity and biologically relevant temperatures. However, high temperature regimes reduce the multiscale simulation performance due to the loss of time scale separation between coarse- and atomic-scale processes. Depending upon conditions, OP-mediated multiscale simulations are much more efficient than conventional MD simulations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported in part by National Science Foundation through the Collaborative Research in Chemistry Program (Grant No. CHE-082651), NIH National Institute of Biomedical Imaging and Bioengineering (Grant No. R21 EB008951), the METAcyt project, and Indiana University College of Arts and Sciences through the Center for Cell and Virus Theory.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

APPENDIX

Principal Component Analysis for Dimensionality Reduction

Principal component analysis (PCA) involves diagonalization of the positional covariance matrix for selected atoms (like C_α in protein backbone)

$$c_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle, \quad i,j=1,\dots,3N \qquad \text{(A1)}$$

where x_i and x_j are atomic coordinates and the $\langle \cdots \rangle$ denote trajectory averages over a selected time window t_l with $l = 1,\dots, L$. This generates an orthogonal set of eigenvectors (i.e., basis vectors or modes), denoted U, each associated with an eigenvalue that indicates the amplitude of fluctuations along that eigenvector. Eigenvalues divided by their sum describe relative contributions of the associated eigenvectors to major conformational changes observed in the trajectory. For reduced representation of the system, a subset of PCA eigenvectors with largest eigenvalues is chosen and the number of these eigenvectors M is normally much smaller than the total number of atoms N,[21,72] i.e.,$U \equiv \{U_k, k = 1,\dots, M\}$ for $M \ll N$.

Simple changes in the definition of space warping OPs (eq 1) enable its construction using the PCA basis vectors. To achieve this, space warping OPs are recast in terms of displacements from a mean configuration. This yields

$$x_i - x_i^0 = \sum_{k=1}^{M} \phi_k U_k + \sigma_i, \quad i=1,\dots,3N \qquad \text{(A2)}$$

where x_i^0 are atomic coordinates of the system reference configuration which can be starting structure of a time window and φ_k are the OPs tracking displacement from this structure. Similarly, the Legendre polynomials constructing u_{ki} are defined as functions of atomic displacements, $\Delta x_i^{t_l}$ ($i = 1,\dots, 3N$) from the selected time window t_l (section II.B). Introducing these u_{ki} into eq 1 yields analogues of A2 in the Legendre basis

$$x_i - x_i^0 = \sum_{k=1}^{M} \phi_k u_{ki} + \sigma_i \qquad \text{(A3)}$$

where the factor u_{ki} is constructed using functions of the form $u_k(\Delta x_i^{t_l})$. This treatment of the Legendre polynomials enables a fair comparison with PCA modes in the context of constructing basis vectors as functions of atomic displacements.

References

1. Xie A, van der Meer AF, Austin RH. Phys Rev Lett. 2002; 88(1):018102. [PubMed: 11800992]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

2. Donko Z, Kalman GJ, Hartmann P, Golden KI, Kutasi K. Phys Rev Lett. 2003; 90(22):226804. [PubMed: 12857333]
3. Delarue M, Dumas P. Proc Natl Acad Sci USA. 2004; 101(18):6957–62. [PubMed: 15096585]
4. Canady MA, Tsuruta H, Johnson JE. J Mol Biol. 2001; 311:803–814. [PubMed: 11518532]
5. Klepeis JL, Lindorff-Larsen K, Dror RO, Shaw DE. Curr Opin Struct Biol. 2009; 19(2):120–7. [PubMed: 19361980]
6. Freddolino PL, Arkhipov AS, Larson SB, McPherson A, Schulten K. Structure. 2006; 14(3):437–449. [PubMed: 16531228]
7. Abraham FF, Walkup R, Gao H, Duchaineau M, Diaz De La Rubia T, Seager M. Proc Natl Acad Sci USA. 2002; 99:5777–5782. [PubMed: 16578876]
8. Ahmed S, Islam S, Mohammed S. IEEE Trans Electron Devices. 2010; 57:164–173.
9. Lange OF, Grubmuller H. J Chem Phys. 2006; 124(21):214903. [PubMed: 16774438]
10. Dey BK, Rabitz H, Askar A. J Chem Phys. 2003; 119(11):5379–5387.
11. Kitao A, Go N. Curr Opin Struct Biol. 1999; 9(2):164–169. [PubMed: 10322205]
12. Elezgaray J, Sanejouand YH. Biopolymers. 1998; 46(7):493–501. [PubMed: 9838873]
13. Amadei A, Linssen ABM, Berendsen HJC. Proteins: Struct, Funct Bioinf. 1993; 17(4):412–425.
14. Bellesia G, Jewett AI, Shea JE. Protein Sci. 2011; 20(5):818–826. [PubMed: 21344535]
15. Arkhipov A, Freddolino PL, Schulten K. Structure. 2006; 14(12):1767–77. [PubMed: 17161367]
16. Freddolino, PL.; Arkhipov, A.; Shih, AY.; Yin, Y.; Chen, Z.; Schulten, K. Application of residue-based and shape-based coarse graining to biomolecular simulations. Chapman and Hall/CRC Press, Taylor and Francis Group; London: 2008. p. 299-315.Chapter 20
17. Gohlke H, Thorpe MF. Biophys J. 2006; 91:2115–2120. [PubMed: 16815893]
18. Maher, M.; Puget, JF.; Backofen, R. Proceedings of the 4th International Conference on Principles and Practice of Constraint Programming. Vol. 1520. Springer-Verlag; Weinheim, Germany: 1998. Constraint techniques forsolving the protein structure prediction problem; p. 72-86.
19. Shreif Z, Ortoleva P. J Stat Phys. 2008; 130:669–685.
20. Riccardi L, Nguyen PH, Stock G. J Phys Chem B. 2009; 113:16660. [PubMed: 20028141]
21. Hayward S, Kitao A, Go N. Proteins: Struct, Funct Genet. 1995; 23(2):177–186. [PubMed: 8592699]
22. Sweet CR, Petrone P, Pande VS, Izaguirre JA. J Chem Phys. 2008; 128:145101. [PubMed: 18412479]
23. Bahar I, Rader AJ. Curr Opin Struct Biol. 2005; 15(5):586–592. [PubMed: 16143512]
24. Chang C-E, Shen T, Trylska J, Tozzini V, McCammon JA. Biophys J. 2006; 90:3880–3885. [PubMed: 16533835]
25. Ayton GS, Voth GA. J Struct Biol. 2007; 157:570–578. [PubMed: 17134912]
26. Pasquali S, Derreumaux P. J Phys Chem B. 2010; 114(37):11957–11966. [PubMed: 20795690]
27. Groot, BLd; Vriend, G.; Berendsen, HJC. J Mol Biol. 1999; 286:1241–1249. [PubMed: 10047494]
28. Nguyen HD, Reddy VS, Brooks CL. J Am Chem Soc. 2009; 131:2606–2614. [PubMed: 19199626]
29. Miao Y, Ortoleva P. J Chem Phys. 2006; 125(4):44901. [PubMed: 16942186]
30. Miao Y, Ortoleva PJ. J Chem Phys. 2006; 125(21):214901. [PubMed: 17166043]
31. Pankavich S, Shreif Z, Miao Y, Ortoleva P. J Chem Phys. 2009; 130:194115. [PubMed: 19466829]
32. Miao Y, Ortoleva PJ. J Comput Chem. 2008; 30(3):423–437. [PubMed: 18636559]
33. Singharoy A, Cheluvaraja S, Ortoleva PJ. J Chem Phys. 2011; 134(4):044104. [PubMed: 21280684]
34. Cheluvaraja S, Ortoleva P. J Chem Phys. 2010; 132(7):075102. [PubMed: 20170252]
35. Shreif Z, Adhangale P, Cheluvaraja S, Perera R, Kuhn R, Ortoleva P. Sci Model Sim. 2008; 15(1–3):363–380.
36. Miao Y, Ortoleva PJ. Biopolymers. 2009; 93(1):61–73. [PubMed: 19728362]
37. Miao Y, Johnson JE, Ortoleva PJ. J Phys Chem B. 2010; 114(34):11181–11195. [PubMed: 20695471]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

38. Joshi H, Singharoy A, Sereda YV, Cheluvaraja SC, Ortoleva PJ. Prog Biophys Mol Biol. 2011; 107(1):200–17. [PubMed: 21802438]
39. Jaqaman K, Ortoleva PJ. J Comput Chem. 2002; 23:484–491. [PubMed: 11908085]
40. Hill, TL. Statistical Mechanics: Principles and Selected Applications. Dover Publications; Mineola, NY: 1987.
41. Iyengar SS, Ortoleva P. J Chem Phys. 2008; 128(16):164716. [PubMed: 18447488]
42. Shea JE, Oppenheim I. J Phys Chem B. 1996; 100:19035e19042.
43. Cukier RI, Deutch JM. J Chem Phys. 1969; 50:36e41.
44. Shreif Z, Pankavich S, Ortoleva PJ. Phys Rev E. 2009; 80(3 Pt 1):031703.
45. Rothman DH, Zaleski S. Rev Mod Phys. 1994; 66:1417.
46. Singharoy A, Sereda YV, Ortoleva PJ. J Chem Theor Comput. 2012 in review.
47. McQuarrie, DA. Statistical Mechanics. Harper and Row; New York: 1976.
48. Sun J, DuFort C, Daniel M-C, Murali A, Chen C, Gopinath K, Stein B, De M, Rotello VM, Holzenburg A, Kao CC, Dragnea B. Proc Natl Acad Sci. 2007; 104(4):1354–1359. [PubMed: 17227841]
49. Ma, SK. Modern Theory of Critical Phenomena (Frontiers in Physics). Perseus Books (Sd); New York: 1976.
50. Whitelam S, Feng EH, Hagan MF, Geissler PL. Soft Matter. 2009; 5:1251–1262. [PubMed: 23227104]
51. Barr, AH. Proceedings of the 11th annual conference on Computer graphics and interactive techniques. ACM; New York: 1984. Global and local deformations of solid primitives; p. 21-30.
52. Pankavich S, Miao Y, Ortoleva J, Shreif Z, Ortoleva P. J Chem Phys. 2008; 128(23):234908. [PubMed: 18570529]
53. MacKerell AD. Abstr Papers Am Chem Soc. 1998; 216:U696–U696.
54. Ayton GS, Noid WG, Voth GA. Curr Opin Struct Biol. 2007; 17:192–198. [PubMed: 17383173]
55. Levy RM, Srinivasan AR, Olson WK, McCammon JA. Biopolymers. 1984; 23(6):1099–1112. [PubMed: 6733249]
56. Liu HJ, Qu CX, Johnson JE, Case DA. J Struct Biol. 2003; 142(3):356–363. [PubMed: 12781662]
57. Speir JA, Munshi S, Wang GJ, Baker TS, Johnson JE. Structure. 1995; 3(1):63–78. [PubMed: 7743132]
58. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. J Comput Chem. 2005; 26:1781–1802. [PubMed: 16222654]
59. Nicholas MG. J Comput Chem. 2006; 27(14):1765–1768. [PubMed: 16917862]
60. Yousuf M, Sahu PC, Jajoo HK, Rajagopalan S, Rajan KG. J Phys F: Met Phys. 1986; 16(3):373.
61. Morellon L, Blasco J, Algarabel PA, Ibarra MR. Phys Rev B. 2000; 62:1022–1026.
62. Schneemann A. Annu Rev Microbiol. 2006; 60:51. [PubMed: 16704342]
63. Day J, Kuznetsov YG, Larson SB, Greenwood A, McPherson A. Biophys J. 2001; 80:2364. [PubMed: 11325736]
64. Singharoy A, Yesnik A, Ortoleva P. J Chem Phys. 2010; 132:1741120.
65. Villa A, Widjajakusuma E, Stock G. J Phys Chem B. 2008; 112(1):134–42. [PubMed: 18069816]
66. García A, Onuchic J. Proc Natl Acad Sci US A. 2003; 100(24):13898–903.
67. Hess B, van der Vegt NFA. J Phys Chem B. 2006; 110(35):17616–17626. [PubMed: 16942107]
68. Antao VP, Lai SY, Tinoco IJ. Nucleic Acids Res. 1991; 19(21):5901–5905. [PubMed: 1719483]
69. Lin MM, Meinhold L, Shorokhov D, Zewail AH. Phys Chem Chem Phys. 2008; 10(29):4227–4239. [PubMed: 18633543]
70. Sorin EJ, Rhee YM, Nakatani BJ, Pande VS. Biophys J. 2003; 85(2):790–803. [PubMed: 12885628]
71. Flint, JS.; Racaniello, VR.; Krug, R.; Flint, SJ.; Enquist, LW.; Skalka, AM.; Racaniello, VR.; Krug, RM. Principles of Virology: Molecular Biology, Pathogenesis, and Control. American Society Microbiology; Washington, DC: 1999.
72. Bradley MJ, Chivers PT, Baker NA. J Mol Biol. 2008; 378(5):1155–1173. [PubMed: 18433769]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

73. Duarte C, Wadley L, Pyle A. Nucleic Acids Res. 2003; 31:4755–4761. [PubMed: 12907716]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

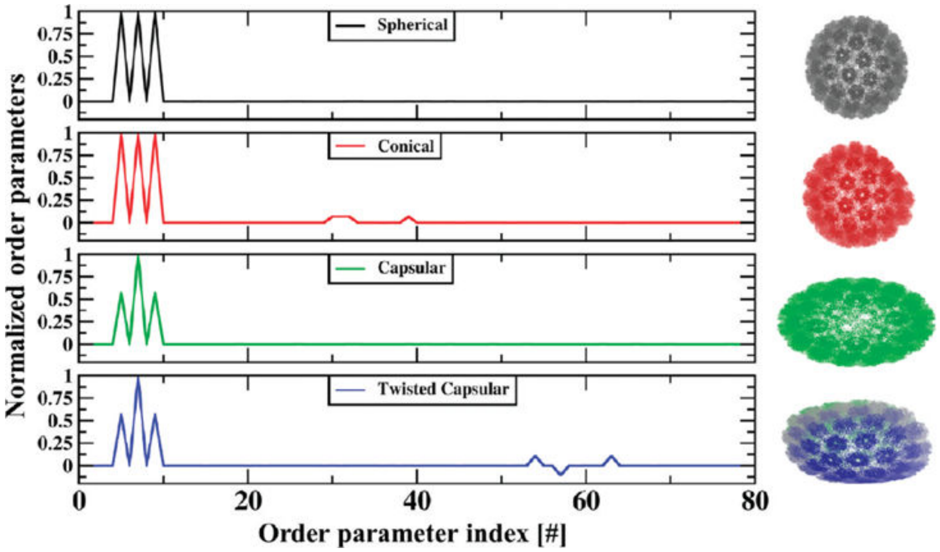


Figure 1.

Normalized space warping OPs with specified k-indices are shown to deform a $T = 7$ capsid into conical, capsular, and twisted capsular forms. Cartesian components of only 3^3 OPs are chosen for this analysis as they are sufficient to capture these transformations.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

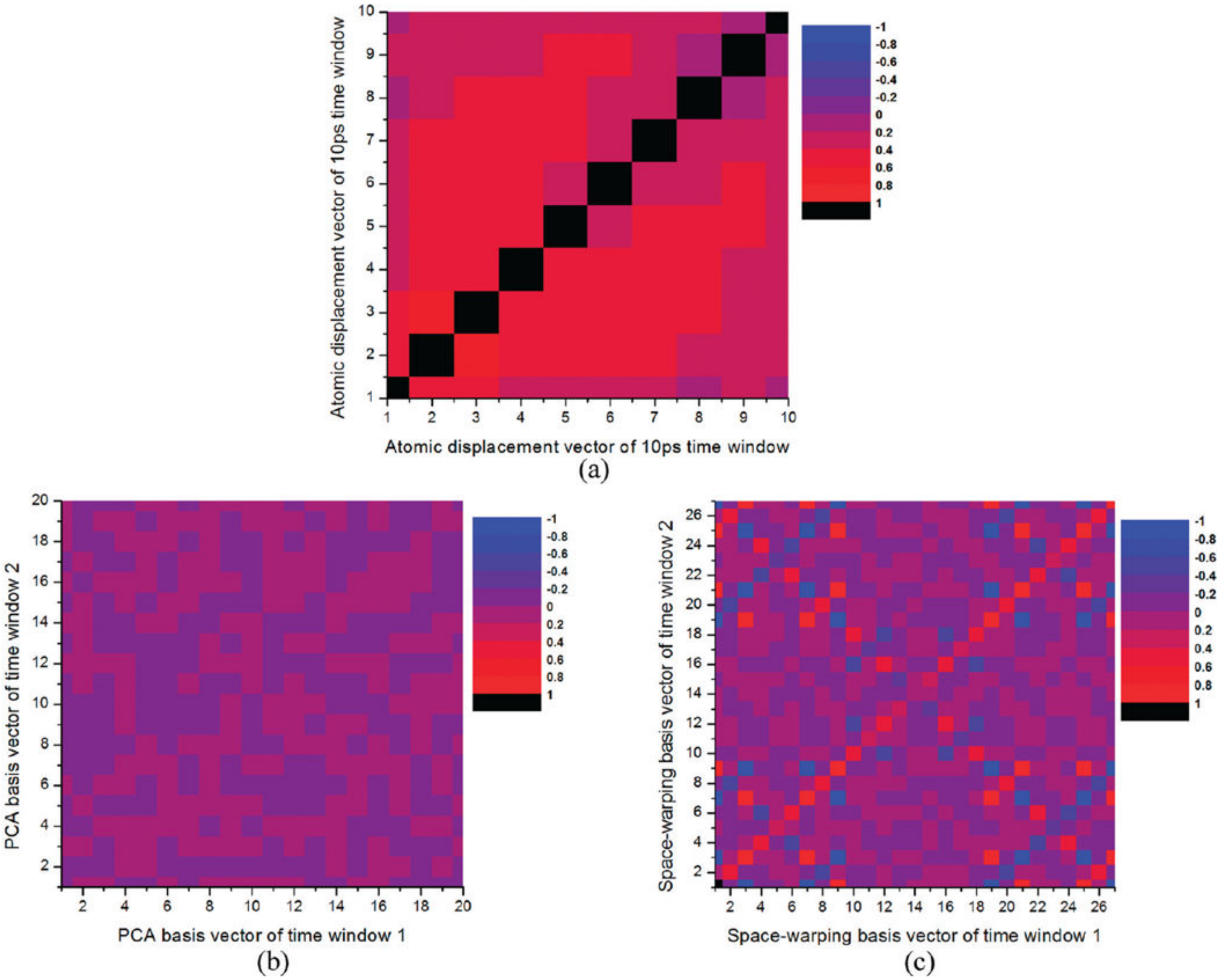


Figure 2.

Correlation matrices of (a) atomic displacement vectors of the ten consecutive time windows of 100 ps ensemble-averaged trajectory and (b) PCA and (c) space warping basis vectors from the first two consecutive time windows. Twenty PCA and 27 Legendre basis vectors are used as they keep the residuals σ_l in eqs A2 and A3 small.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

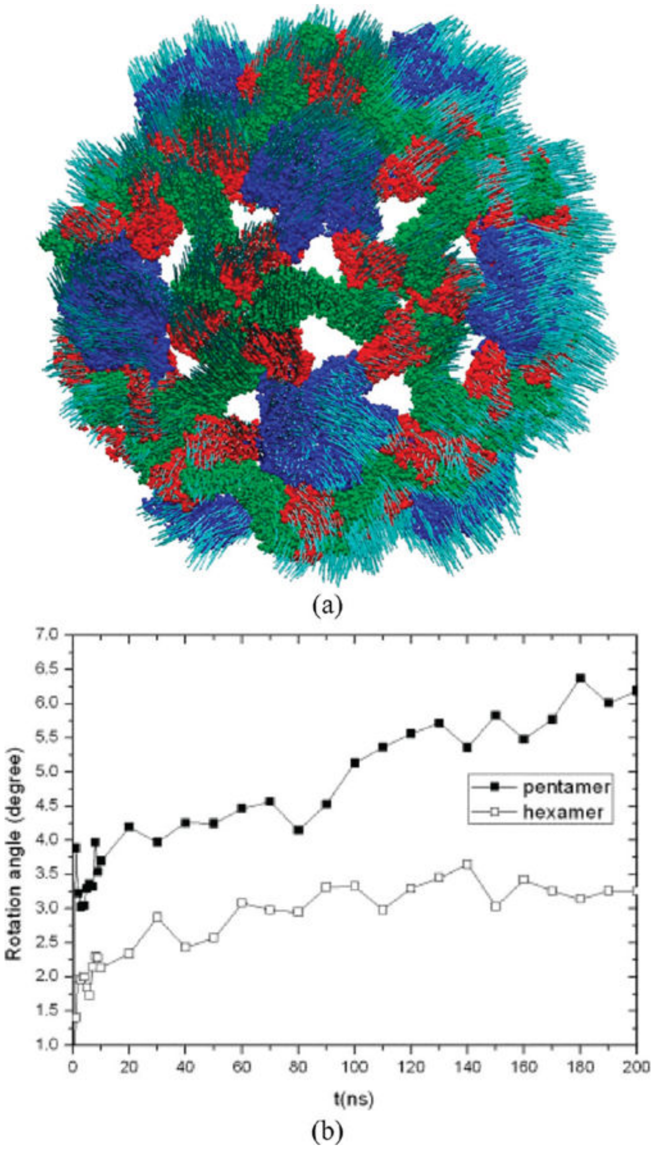


Figure 3.

200 ns MD/OPX simulation of the shrinkage of swollen CCMV capsid: (a) CPK representation of C_α atoms of the resulting structure with arrows indicating the atomic displacements from their original positions in the starting PDB structure to their final positions and (b) time courses of the average rotation angle for pentamers and hexamers calculated via superimposing their structures to the initial configurations.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

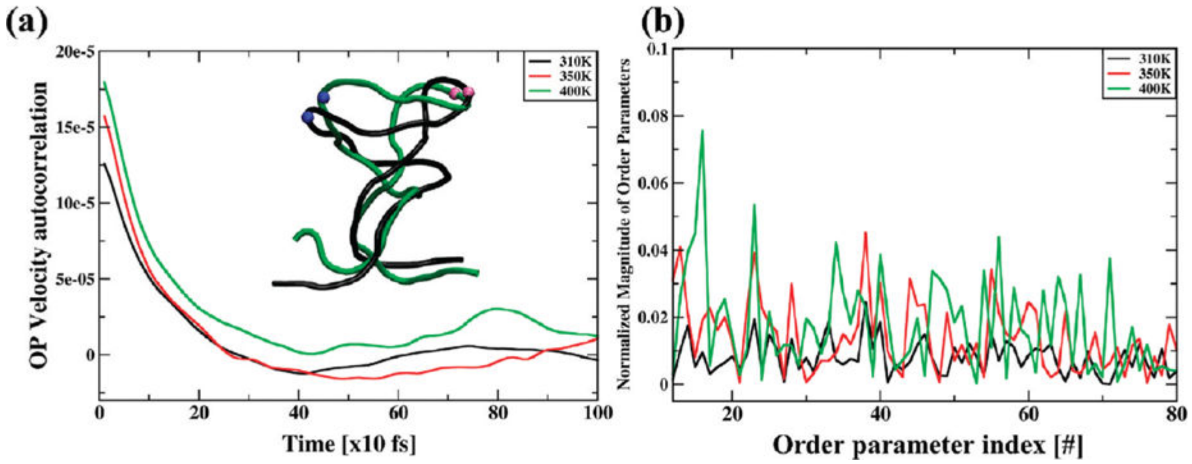


Figure 4.

(a) Velocity autocorrelation functions for a typical OP (Φ_{100X}). Absence of a long-time tail indicates lack of coupling to other slow variables not included in the set of OPs. As temperature increases, so does the area under a curve, implying greater OP diffusion. (b) Appearance of new OPs ($\underline{k} = 2$ and 3 here) indicates emergence of collective motions as temperature increases. Some of these OPs correspond to those for twisting and tapering (Figure 1), thereby capturing the emergence of nonlinear deformation about thermo-labile centers of the RNA helix (inset). These emergent OPs are obtained from the population of growing residuals as described in Figure S3 in the Supporting Information.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

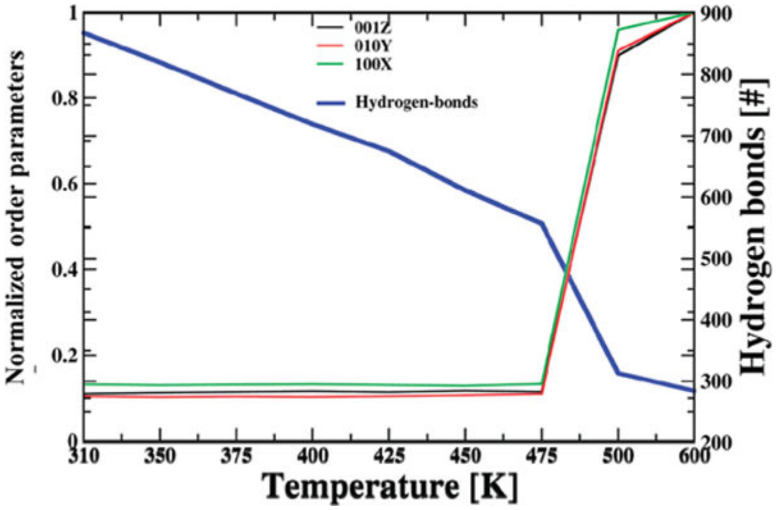


Figure 5.

Phase transition behavior of STMV RNA captured via the quasi-discontinuous change of space warping OPs between 475 and 500 K suggesting that these OPs are appropriate indictors of a change in organization even if the nature of symmetry breaking is not apparent. The change in OPs indicates thermal expansion of the RNA. Similar behavior is reflected in the average number of intranucleic acid hydrogen bonds that rapidly decreases across this region.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

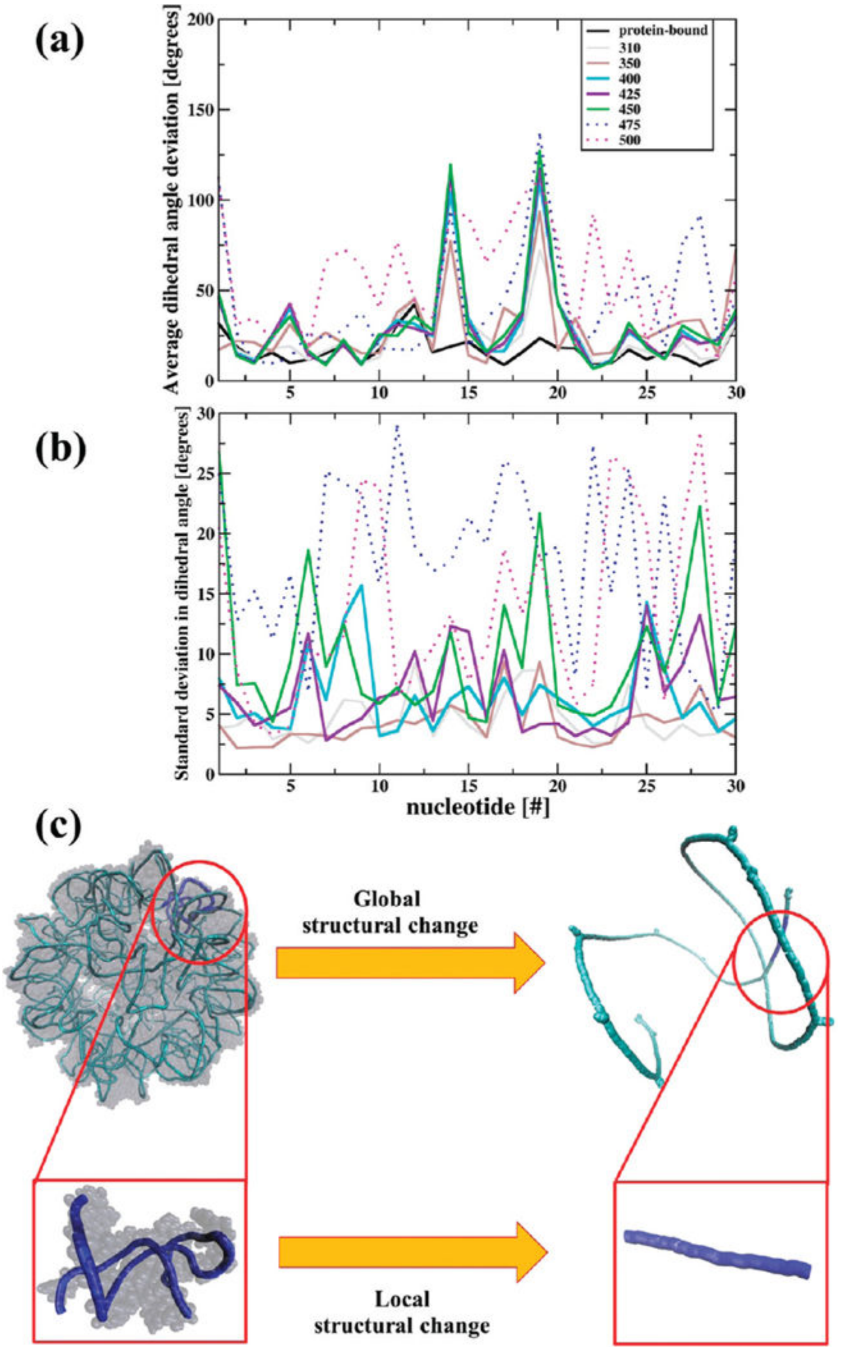


Figure 6.

(a) Average deviation of nucleotide dihedral angles from their initial values at $t = 0$ for 30 nucleotides comprising of the RNA helix for temperatures between 310 and 500 K. At temperatures lower than the transition temperatures, nucleotides 14 and 19 show maximum deviation from their room temperature values. At temperatures in the transition zone, a more RNA-wide motion occurs, as indicated by the increase in dihedral deviation of neighboring nucleotides. In contrast, the protein bound RNA helix is the least mobile. (b) rms fluctuations about mean dihedral angles showing distinct increase with temperature.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Dihedral angles are calculated using the AMIGOS[73] program. Only η-type dihedrals[73] are plotted for illustrative purposes. At each temperature, the ensemble average dihedrals are averaged over all the thirty RNA helices to obtain data for this plot. (c) Global and local changes in the structure of viral RNA accompanying the transition between states of uncertain symmetry.