JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Coarse-Grained and Atomistic Simulations of the Salt-Stable Cowpea Chlorotic Mottle Virus (SS-CCMV) Subunit 26−49: β-Barrel Stability of the Hexamer and Pentamer Geometries

Tristan Bereau,*[,†,§] Christoph Globisch,[‡] Markus Deserno,[†] and Christine Peter[‡]

[†]Department of Physics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, United States

[‡]Max-Planck-Institute for Polymer Research, Ackermannweg 10, D-55128 Mainz, Germany

ABSTRACT: A combination of coarse-grained (CG) and atomistic simulations provides a suitable computational framework to study unstructured regions of proteins, for which experimental data are often lacking or limited. In this work, we combine CG and atomistic simulations with clustering algorithms and free energy reweighting methods to explore the conformational equilibrium of certain regions of the salt-stable cowpea chlorotic mottle virus (SS-CCMV). In particular, we focus on the geometry of converging strands (residues 26−49) from contacting subunits at the 3-fold (hexamer) and 5-fold (pentamer) symmetry points of the capsid. We show the following: (*i*) The simulations reproduce the experimentally observed β-barrel for the hexamer. (*ii*) The pentamer geometry is unable to stabilize a β-barrel conformation; it assumes various states instead, again in accordance with the experimental results which do not indicate a well-defined structure for the pentameric interface. (*iii*) Atomistic simulations of the backmapped CG structures remain relatively stable, indicative of plausible CG conformations and slow kinetics on the atomistic level.

1. INTRODUCTION

Cowpea chlorotic mottle virus (CCMV), a member of the Bromoviridae family, forms a model system for understanding protein−protein interactions that dictate icosahedral virus assembly.[1−3] In the present paper, we study a salt-stable CCMV mutant (SS-CCMV), for which the structure was determined from X-ray data.[4]

SS-CCMV's $T = 3$ capsid is made of 180 identical proteins (subunits), which form 12 5-fold and 20 6-fold interfaces.[5] While all proteins have the same amino acid sequence, they adopt slightly different folds due to their positions in symmetrically inequivalent locations. Coat proteins of type A form pentamers, while coat proteins B and C form hexamers (Figure 1). Speir et al.'s crystal structure[4] resolved coat protein A starting from residue 42 and proteins B and C starting from residue 26. The lack of structure for residues 1−25 is attributed to its highly basic sequence, which interacts strongly with the viral RNA packaged in the interior region of the virus particle (see ref 4 and references therein). Next along the sequence, the six converging strands from the six contacting subunits at the 3-fold symmetry axis (i.e., hexamer—proteins B and C) have been associated with a β-barrel for residues 26−35. On the other hand, the five converging strands at the 5-fold symmetry axis (i.e., pentamer−protein A) have not been associated with any structure so far for residues 26−41. This issue forms the basis of this work: can a computational approach help us explain the absence of a β-barrel in the case of coat protein A? Can simulations propose plausible conformations for those regions, where no unique native structure can be resolved experimentally? In particular, we will rely on a combined CG and atomistic molecular dynamics (MD) simulation approach[6,7] that has been shown to be efficient in sampling unstructured regions of proteins (e.g., intrinsically disordered proteins[8]), for which experimental data are often lacking or limited. To this end, the system's free-energy landscape is first probed using a CG model, which allows thorough sampling; the identified minima's structures are then refined by means of backmapped atomistic simulations.[9]

The present work consists of a comparative study of various hexameric and pentameric interfaces at both CG and atomistic resolution. For this, we do not need to study the whole virus but only focus on suitably chosen neighborhoods around the interfaces. In addition to the barrel-forming hexamer[4] (see amino acid sequence in Table 1; "hex"), an unstructured glycine-rich mutant[2] (Table 1; "hex-gly") will also be studied. While much is already known about these two systems, they will allow a validation of the CG model before applying it to the pentameric system ("pent"). Furthermore, the ability to form β-barrels will be assessed by performing a set of control simulations in which the force field is strongly biased in favor of β structures for the hexamer and pentamer—"hex-CGβ" and "pent-CGβ," respectively. The main conformations sampled from these systems will then be backmapped and simulated atomistically ("hex-at" and "pent-at").

The combined CG/atomistic approach taken here will show that: (*i*) the CG model can correctly predict the stabilization of a β-barrel for the hexamer and the glycine-rich mutant's lower stability, (*ii*) the pentamer is unable to stabilize such a barrel conformation and instead assumes a diverse set of disordered states, (*iii*) β-favoring CG simulations suggest a geometric incompatibility between pent-WT and the hydrogen-bond pattern of a β-barrel, and (*iv*) atomistic simulations of backmapped CG structures remain stable over 50 ns, indicative

Special Issue: Wilfred F. van Gunsteren Festschrift

Received: December 13, 2011
Published: April 23, 2012

   dx.doi.org/10.1021/ct200888u | *J. Chem. Theory Comput.* 2012, 8, 3750−3758

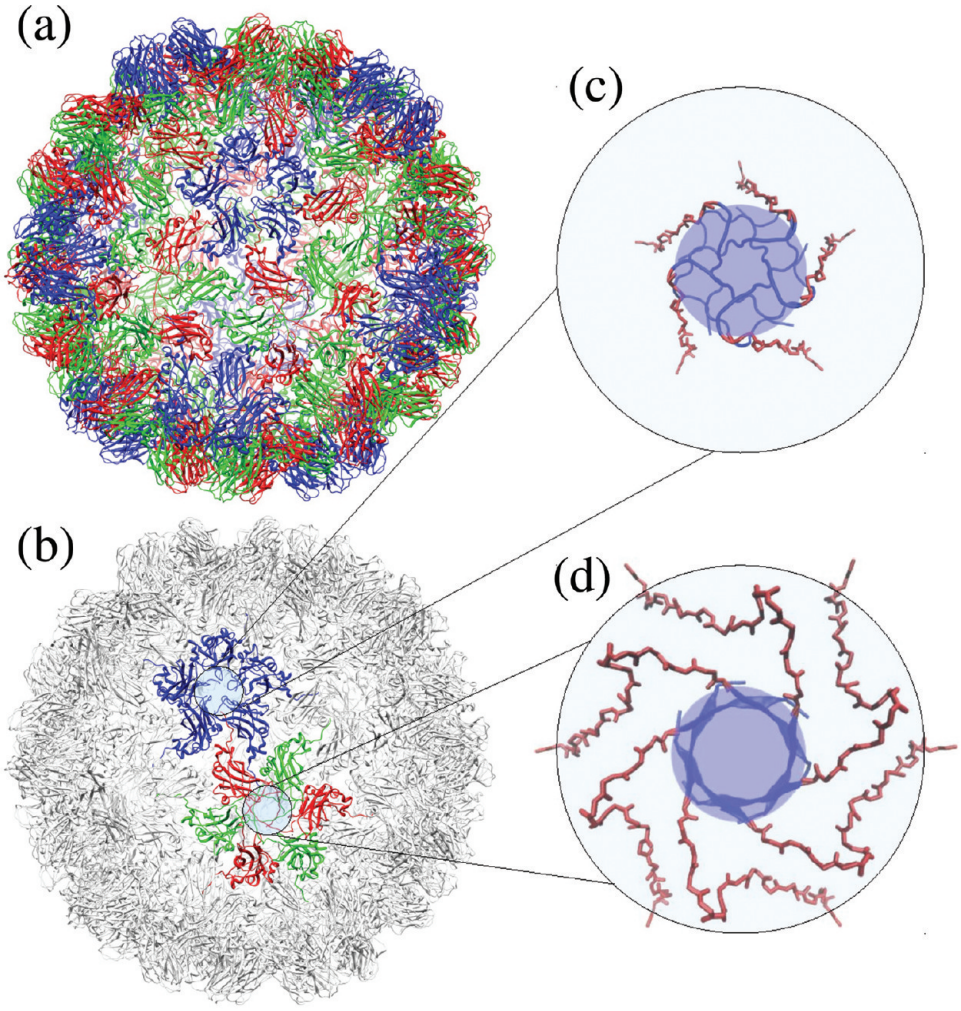

Figure 1. (a) CCMV's viral capsid, formed of three types of coat proteins (color-coded in the figure), which form hexameric and pentameric interfaces. (b) The same capsid with one hexameric and pentameric interface highlighted. Zooming into both types of interfaces: top view, backbone conformation of the initial pentamer (c) and hexamer (d) structures studied in this work, visualized with VMD.[10] All residues initially located within 10 Å from the center of mass of the interface (represented by a filled circle) were free to move; the remaining parts were constrained (restrained) during the entire course of the CG (atomistic) simulation.

Table 1. Amino Acid Sequences of the Different Subunits of SS-CCMV,[4] from Residues R26 to G49, Studied in This Work: Hexamer (hex), Glycine-Rich Hexamer (hex-gly), and Pentamer (pent)[a]

protein	monomer sequence
hex	RVVQ P_{30}VIVE $\underline{P_{35}IASG\ Q_{40}GRAI\ K_{45}AWTG}$
hex-gly	RVVQ G_{30}GGGG $\underline{G_{35}IASG\ Q_{40}GRAI\ K_{45}AWTG}$
pent	RVVQ P_{30}VIVE P_{35}IASG $\underline{Q_{40}GRAI\ K_{45}AWTG}$

[a]Underlined amino acids were position-constrained during the CG (restrained in atomistic) simulations; see also Figure 1.

of realistic initial conformations and slow kinetics. Overall, the CG model is successfully used to suggest a set of low free-energy conformations for these flexible, partly unstructured regions.

2. METHODS

2.1. Initial Conformations. Initial conformations used in this work are based on Speir et al.'s crystal structure[4] (PDB: 1ZA7). Both hexamer and pentamer geometries were studied between residues 26 and 49. Because proteins B and C have been resolved starting from residue 26, extracting the hexamer's initial conformation was straightforward. In the pentamer case, the missing structure was generated by a homology model based on the structure of the Brome Mosaic Virus[11] (BMV; PDB: 1JS9). Indeed, BMV and (SS-)CCMV belong to the same family of Bromoviridae and share a 70% sequence identity.[12] Moreover, there is good agreement between SS-CCMV's chains B and C and BMV's chain C. Chain A's missing residues were here reconstructed using the comparative protein modeling program MODELLER 9v7.[13] The resulting structure—made of five identical subunits—was used as the initial pentameric conformation.

We point out that the free-energy calculations presented below should not depend on the selected initial conformations, considering the extensive amount of sampling obtained during our CG simulations.

2.2. Environment-Mimicking Position Constraints. In this work, we focus on the pentameric (hexameric) interface and constrain the rest of the system to mimic its stability. To do so, the five (six) chains that form the interface are modeled from residues 26 to 49, and we prevent a number of residues from the end of each chain from moving. The number of

dx.doi.org/10.1021/ct200888u | J. Chem. Theory Comput. 2012, 8, 3750−3758

constrained residues was motivated by the following: the pentamer's residue interval 26−40 corresponds to the entire range for which X-ray data are lacking[4] (excluding the basic N-terminal region 1−25 that interacts with the viral RNA); it was thus natural to constrain all residues starting from #41. In order to compare the pentamer with the hexamer, two options prevailed: simulate either the same set of amino acids or the same geometric region from the interface's center. Because of the inherently different geometric arrangements of the chains around the hexameric and pentameric interfaces (see Figure 1), we have found a residue-based constraint criterion (i.e., option #1) to be inappropriate. Instead, we found it best to constrain all residues located further than 10 Å from the center of mass of the interface, which is exactly the distance beyond which we know the pentamer structure. Though a longer distance criterion would have been preferable, the lack of salt bridges and both side-chain−side-chain and side-chain−backbone hydrogen bonds in the CG model prevents the system from stabilizing the correct global native state (which does contain many such interactions close to the interface region in question). As shown in Table 1, our simulations allowed residues 26−40 and 26−34 to evolve freely for the pentamer and hexamer, respectively (see Figure 1).

2.3. Coarse-Grained Simulations. We performed CG simulations based on an intermediate resolution, implicit-solvent peptide model.[14] This model, while offering a reduced (i.e., simplified) chemical resolution, accounts for amino acid specificity and is capable of representing genuine secondary structure without explicitly biasing the force field toward any particular conformation (native or not). This is in contrast with other force fields that require prior knowledge of the protein's structure, e.g., Gō models only assign nonbonded attractive interactions to nearby residue−residue contacts present in the input native conformation.[15] The absence of known pentamer structure in this study clearly motivates the use of a force field that does not rely on any input structure. The present model represents an amino acid by four beads, three of which are devoted to the protein backbone. It was parametrized using a top-down scheme in which phenomenological interactions are incorporated to reproduce different properties of the system (e.g., hydrophobicity, hydrogen bonds). After parametrization, the model was shown to be capable of folding several simple helical proteins and reproduce an oligopeptide aggregation scenario using the same force-field parameters and without any primary sequence dependent ("custom-made") bias.[14] These successful tests demonstrate that the model is both transferable (i.e., the same force field can be applied to different proteins) and capable of reproducing a qualitatively accurate α/β balance. Recently, fine details of cooperative folding transitions for helical peptides were studied with this model[16,17] and were shown to be in agreement with expectations. Unlike α helices, finding systems that permit well-stabilized β structures is rather difficult because single β strands so easily fall apart, and a complete β-barrel (e.g., OmpA[18]) introduces many complications (e.g., sampling difficulties; presence of both water and membrane environments). The CCMV virus barrel, however, stands as an ideal system for the present CG model because of its stability and limited size.

While this model correctly balances different secondary-structure motifs, we have found it useful (in specific cases only) to artificially tweak the force field in order to strongly favor β-sheet conformations, because we can then test against the objection that we might understabilize β structures. This artificial bias was obtained by increasing the strength of a specific interaction—namely, the nearest-neighbor dipole−dipole interaction k_{dip}, which effectively favors β over α conformations (see ref 14 for details). While $k_{dip} = -0.3\varepsilon$ in the original force field (where ε is the intrinsic unit of energy of the CG model), the altered force field consisted of $k_{dip} = -0.85\varepsilon$ (this value offers significant stabilization of β conformations while retaining fast kinetics along the backbone dihedrals). To distinguish simulations run with the original and the (artificial) β-biased force field, they will be suffixed with "-CG" and "-CGβ," respectively (e.g., hex-CGβ is the hexamer run with the biased force field).

We performed replica-exchange MD simulations using the ESPResSo package[19] and ran all simulations in the canonical (NVT) ensemble using a Langevin thermostat with friction constant $\Gamma = \tau^{-1}$, where τ is the intrinsic unit of time of the CG model. The CG unit of energy, ε, relates to thermal energy at room temperature via $\varepsilon = k_B T_{room} \simeq 0.6$ kcal mol^{-1}. The temperature T is expressed in terms of the intrinsic unit of energy $[T] = \varepsilon/k_B$. The equations of motion were integrated with a time step $\delta t = 0.01\tau$. Each temperature in the following set was simulated for a total time of $10^6\tau$ using a parallel tempering scheme: $k_B T = \{0.7, 0.8, ..., 1.4\}\varepsilon$.

A cluster analysis was used to study the structural heterogeneity of the conformations sampled. Clusters were extracted from the *common-nearest-neighbor* algorithm described in Keller et al.[20] The cluster algorithm was applied on a total of 4000 conformations per system—500 snapshots per replica—irrespective of the sampled temperature. Applying the cluster algorithm on all replicas simultaneously, rather than a single temperature of interest, will allow the calculation of relative free energies for all clusters (see below). The root-mean-square distance (RMSD) was used as a metric to quantify separation between conformations. No alignment was performed prior to RMSD calculation. The clustering algorithm relies on two parameters, *nndc* (nearest-neighbor distance cutoff) and *nnnc* (number of nearest neighbors cutoff), which were always set to 4 and 10, respectively (for details regarding these parameters, see ref 20). We have found that varying these two parameters within a reasonable interval gave consistent results (data not shown).

Free energy calculations were performed using the weighted histogram analysis method (WHAM).[21−24] WHAM calculates a minimum-variance estimator for the density of states by combining simulations performed at different temperatures. By coupling the above-mentioned cluster algorithm and WHAM, we can (*i*) subdivide the ensemble of sampled conformations into a discrete set (i.e., *clusters*) and (*ii*) quantify their relative stability at a given temperature (set to $T = T_{room}$ throughout this work). Because WHAM relies on a temperature-reweighting procedure, every single snapshot of all replicas must be associated with a cluster—the cluster algorithm must therefore be run over the entire set of replicas prior to calculating free energies. While high-temperature replicas access a larger diversity of conformations, the incorporation of replicas at *low* temperatures (i.e., $T < T_{room}$) increases the statistical weight of low energy structures. This ensures that such conformations contribute significantly to the cluster algorithm, compared to the large diversity of high energy structures accessed by high-temperature replicas. Error bars were evaluated from bootstrapping the raw data.[25]

2.4. Atomistic Simulations. Atomistic simulations were performed with the Gromacs 4.5.4 simulation package,[26] the

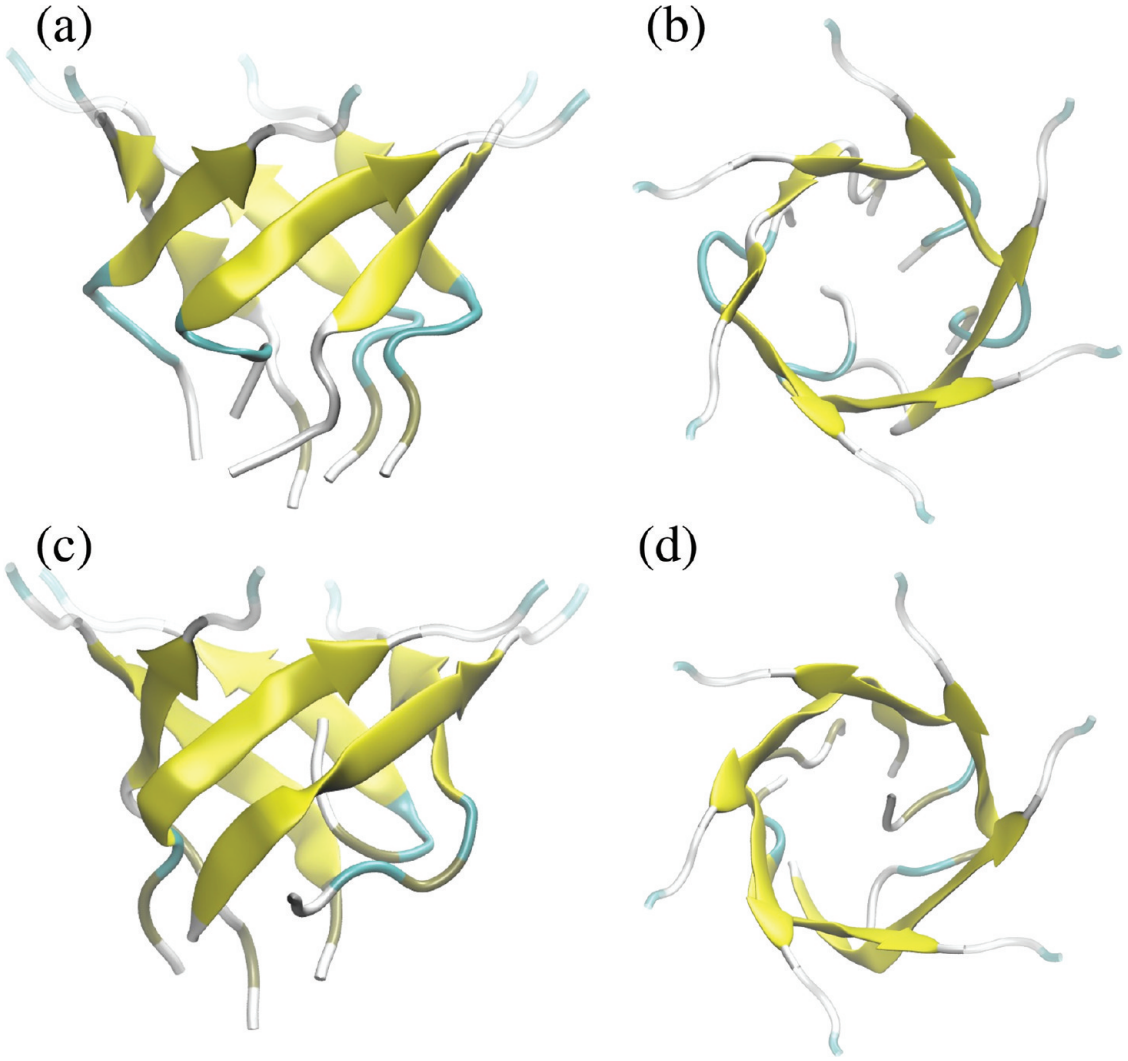

Figure 2. Side and top views of the most populated cluster of hex-CG (a, b) and hex-CGβ (c, d), respectively. Residues 26−37 of each chain are shown. Yellow ribbons correspond to β sheets, as assigned using stride.[33] The β-barrel extends over amino acids 30−34 (see sequence in Table 1).

atomistic GROMOS 54a7 force field,[27] and the SPC water model.[28] Each protein complex was solvated in a rectangular box and neutralized by chloride ions (i.e., 12 and 10 for hexameric and pentameric systems, respectively). The initial box size was set to ensure a distance between the box edge and the protein of at least 1 nm. Simulations were run under constant temperature (T = 298 K) and pressure (P = 1 bar) conditions. During the equilibration phase, the temperature and pressure were maintained close to their target values by means of the Berendsen weak-coupling method[29] with coupling constants τ_T = 0.1 ps and τ_P = 0.5 ps, respectively. For the subsequent production run, the Berendsen weak-coupling method controlled the pressure (τ_P = 5 ps), while the temperature was controlled by a Langevin thermostat with friction constant Γ = 1 ps^{-1}. The nonbonded interactions were calculated with a twin-range cutoff scheme: (*i*) the short-range van der Waals and Coulomb interactions—within a 1.0 nm distance—were evaluated at every time step, and (*ii*) the longer range van der Waals interactions—within 1.4 nm—were updated together with the neighbor list every 10 time steps. The long-range Coulomb interactions were calculated by the PME method[30,31] with a grid spacing of 0.12 nm. A time step of 2 fs was applied throughout, and all bonds were constrained by the LINCS algorithm.[32] Each production run lasted 50 ns.

Initial atomistic conformations were generated by back-mapping representative cluster centers sampled in the CG simulations. The high resolution of the CG model allowed a straightforward reconstruction of the backbone atoms. The coordinates of the side-chain atoms were reconstructed using the MODELLER 9v7 software.[13]

The solvated, neutralized systems were first energy-minimized with position restraints on the protein atoms (force constant: 1000 kJ mol^{-1} nm^{-2}) and subsequently without restraints (for residues within 10 Å of the system's axis; all others were fixed for the entire course of the simulation; see Figure 1) using both a steepest descent method and a conjugate-gradient algorithm. Several 200-ps-long equilibration simulations were performed with a decreasing number of restrained atoms: (*i*) all protein atoms at first, (*ii*) then the backbone atoms, and (*iii*) only the α carbons (i.e., C$_\alpha$).

3. RESULTS

3.1. Unbiased CG Simulations. *3.1.1. Hexamer (hex-CG).* We first ran CG simulations of the hexamer system followed by a clustering of the sampled conformations. The most populated

cluster of hex consists of a complete β-barrel, as shown in Figure 2a and b. The secondary structure assignment algorithm stride recognized β conformations in residues 30−34 of all chains but one, for which one residue was identified as coil. While this defect could be a specific feature of the cluster, we believe that it is most likely the result of a fluctuation which pushed the local chain conformation beyond the threshold set by STRIDE for β conformations.

As a means to evaluate the heterogeneity of conformations sampled during the simulation, Table 2 shows the relative

Table 2. Proportion of Conformations (i.e., Number of Elements—Irrespective of the Sampled Temperature—Divided by the Total Number of Conformations) Assigned to the 10 Most Populated Clusters for hex-CG, hex-gly-CG, pent-CG (top), hex-CGβ, and pent-CGβ (bottom)[a]

cluster	hex-CG	hex-gly-CG	pent-CG
#1	52%	9.2%	8.2%
#2	12%	7.3%	7.9%
#3	9.3%	5.7%	5.6%
#4	<1%	3.0%	5.3%
#5	<1%	2.8%	5.1%
#6	<1%	1.8%	3.0%
#7	<1%	1.6%	2.8%
#8	<1%	1.4%	1.6%
#9	<1%	1.1%	1.5%
#10	<1%	1.0%	1.4%
cluster	hex-CGβ		pent-CGβ
#1	84%		12%
#2	1.4%		11%
#3	<1%		10%
#4	<1%		9.4%
#5	<1%		5.0%
#6	<1%		4.3%
#7	<1%		2.7%
#8	<1%		2.4%
#9	<1%		2.2%
#10	<1%		<1%

[a]Because these are averaged over all replicas, they do not represent canonical weights; instead, they illustrate how many clusters represent most conformations sampled.

weight of the 10 most populated clusters, irrespective of the temperature at which any conformation was sampled. While this observable certainly does *not* quantify canonical populations (i.e., cluster stability), it shows that a few clusters are sufficient to represent most conformations sampled. Clearly, hex-CG's cluster #1 (Figure 2a and b) is overwhelmingly sampled, suggestive of a strong stability at low temperatures.

To quantify the relative stability (i.e., canonical weight) of the most populated clusters, we calculate their free energies by reweighting the contributing structures using WHAM. Figure 3 shows the relative free energy of each cluster calculated at $T = T_{room}$. $F = 0k_BT_{room}$ was set by the most stable (i.e., lowest free energy) cluster sampled. The horizontal line corresponds to $F = k_BT_{room}$. Clusters within this free energy difference from the most stable cluster will also be sampled significantly around room temperature, and hence their number provides an indication on the structural variability of conformations thermally accessible at $T = T_{room}$: more clusters implies a larger diversity of representative structures and, therefore, a reduced stability of any particular one. In the case of hex-CG

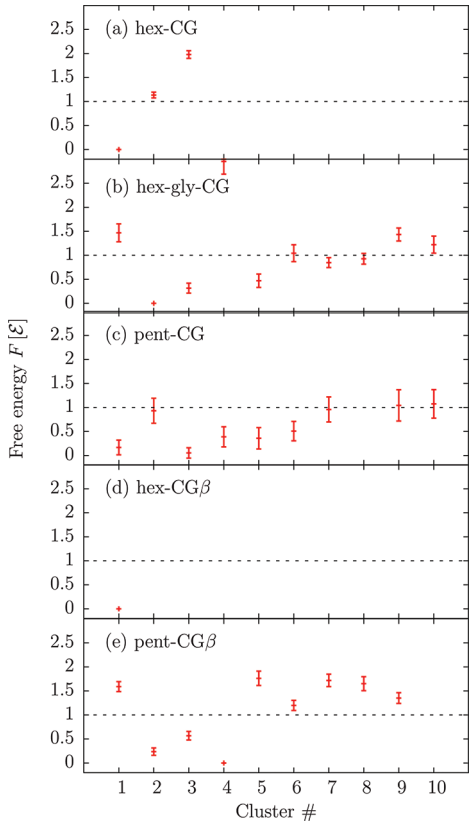

Figure 3. Free energy F calculated at T_{room} for the 10 most populated clusters of (a) hex-CG, (b) hex-gly-CG, (c) pent-CG, (d) hex-CGβ, and (e) pent-CGβ. Horizontal dashed lines indicate $F = k_BT_{room}$. Missing data points correspond to $F > 3k_BT_{room}$.

(Figure 3a), cluster #1 (Figure 2a and b) is virtually the only type of thermally accessible conformation (i.e., only data point below the $F = k_BT_{room}$ line). This indicates that the β-barrel is the most stable structure with no competing conformation, in agreement with Speir et al.'s X-ray data.[4]

3.1.2. Gly-Rich Hexamer (hex-gly-CG). To check that the CG model is capable of reproducing important structural changes observed in mutation experiments, we now turn to CG simulations of the glycine-rich hexamer mutant hex-gly-CG (sequence shown in Table 1). Compared to hex-CG, the larger number of clusters required to represent most conformations (independently of temperature) is suggestive of a more heterogeneous population (Table 2): more clusters are significantly populated, and none of them overwhelms the others. This variability is canonically quantified in the relative free energy of stability at T_{room} of the most populated clusters (Figure 3b), which indicates that five out of ten are within the $F = k_BT_{room}$ threshold. Snapshots of the three most stable clusters (#2, #3, and #5) show no barrel geometry, even though a large number of β sheets are present (Figure 4a−c). Glycines' reduced steric constraints[34] permit a structurally diverse ensemble of favorable conformations. In addition, the absence of specific interactions (e.g., salt bridge, side-chain hydrogen bond)—some of which are *not* modeled by the CG force field—contributes to the lack of stability of hex-gly-CG's clusters. The results for this glycine-rich mutant also demonstrate that the environment-mimicking position constraints do not inevitably bias toward a barrel formation.

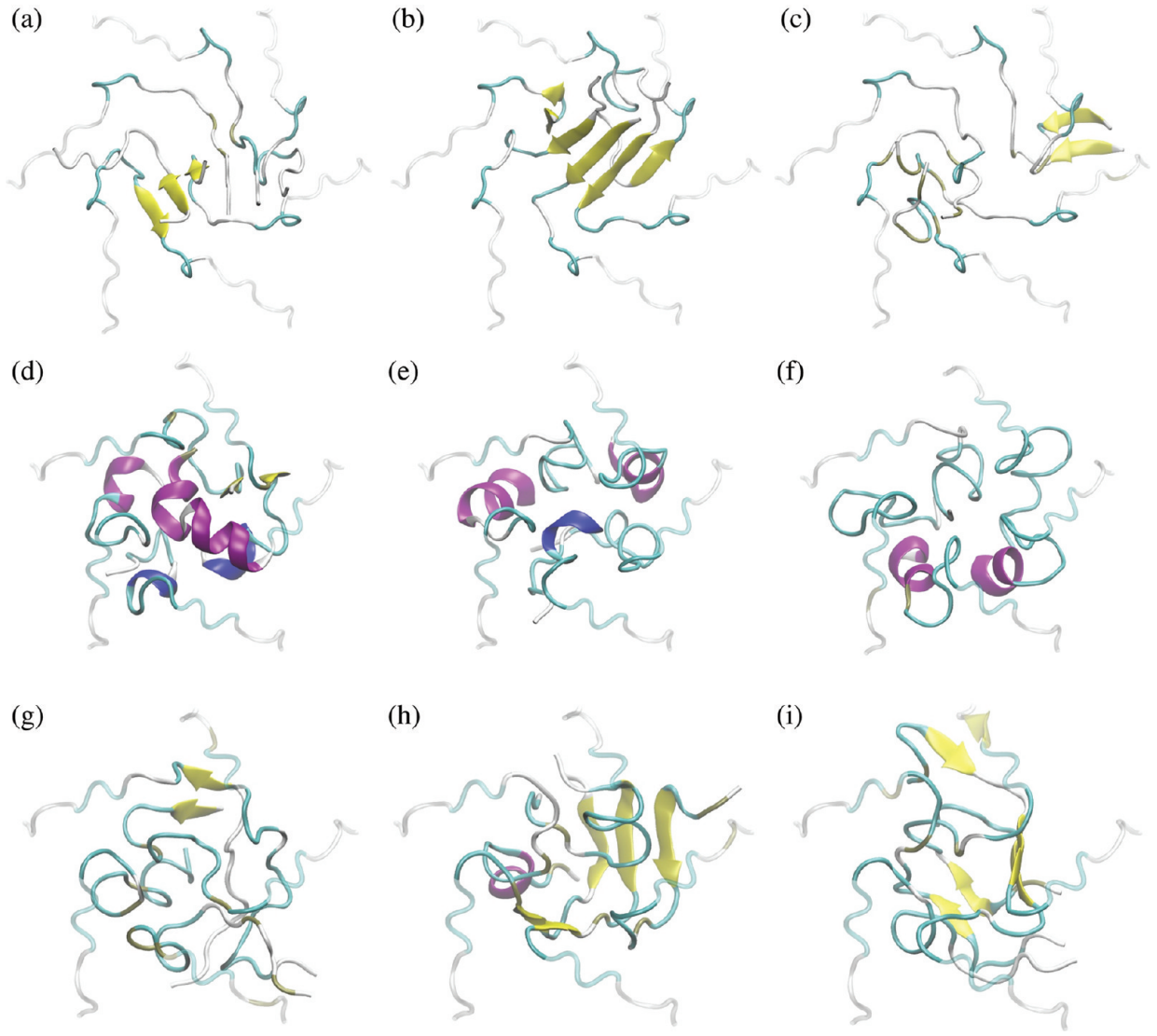

Figure 4. Top views of the three most stable clusters at T_{room} of hex-gly-CG: (a) #2, (b) #3, and (c) #5 (see Figure 3). pent-CG: (d) #1, (e) #3, and (f) #5. pent-CGβ: (g) #2, (h) #3, and (i) #4 (see Figure 3). Residues 26−48 of each chain are shown.

3.1.3. Pentamer (pent-CG). Similar to hex-gly, CG simulations of the pentamer (pent-CG) show a large number of sampled clusters over the entire range of temperatures (Table 2). The free energy at $T = T_{room}$ shows that almost all of the 10 most populated clusters are stable within the $1k_BT_{room}$ threshold (Figure 3c). Figure 4d−f illustrates the conformations of the three most stable clusters. Two points are worth mentioning: (*i*) no conformation resembles the formation of a β-barrel, and (*ii*) there are significantly more helices than in the conformations of hex-gly (Figure 4a−c). Overall, the results suggest a strong variability of conformations, in agreement with the absence of discernible structure in the X-ray data within this region of the protein.[4] Considering that hydrogen-bond formation is a strong driving force in protein stability, the absence of any β-sheet structure raises the question of whether this pentamer geometry is at all compatible with the stabilization of β strands. The following paragraphs address this question.

3.2. β-Favoring Biased CG Simulations. While the conclusions presented above agree with the X-ray results obtained for the hexamer and pentamer,[4] they do not explain why the two systems behave so differently. One possible explanation for the absence of a β-barrel in the pentamer system is a geometric incompatibility: the known arrangement of chains leading into the 5-fold symmetry point is not conducive to the hydrogen-bond pattern required for a stable barrel.

To test this, we alter the CG force field to strongly bias conformations in favor of β regions of the Ramachandran plot. As a proof of concept, the biased force field was first applied to hex (denoted hex-CGβ) to demonstrate its ability to increase β stability. Next, the pentamer system was studied in the same setting ("pent-CGβ") to distinguish whether β-barrels are absent in the pentamer because they are (*i*) *thermodynamically* or (*ii*) *geometrically* unfavorable.

3.2.1. Hexamer (hex-CGβ). Biased CG simulations of hex show that most conformations sampled—irrespective of the temperature—were grouped in one large cluster (Table 2). This cluster is represented in Figure 2c and d: a complete β-barrel extending over the same sequence as hex-CG (Figure 2a and b). Its stability is strongly increased compared to hex-CG, as can be seen in the free energy profile (Figure 3d): the second most stable cluster has a free energy $F > 3k_BT_{room}$. Clearly, this β-barrel is much more stable than hex-CG's unbiased

dx.doi.org/10.1021/ct200888u | J. Chem. Theory Comput. 2012, 8, 3750−3758

simulation; it therefore demonstrates the altered force field's preference for β structures.

3.2.2. Pentamer (pent-CGβ). Unlike hex-CGβ, biased CG simulations of the pentamer (pent-CGβ) yield a highly heterogeneous population of clusters (Table 2) and a rather large variability of canonically stable conformations at $T = T_{room}$ (Figure 3e). The conformations of the three most populated clusters (Figure 4g–i) comprise many more β sheets than pent-CG (Figure 4d–f) but do not form any β-barrel. Considering the force field's strong bias in favor of β conformations, the results suggest that the inability of the chains around the 5-fold axis to form a β-barrel is not merely due to unfavorable *energetics* but rather an incompatibility arising form the systems' *geometry*.

Using simple theoretical calculations,[35,36] the stability difference between pentameric and hexameric barrels appeared to be borderline. For this reason, we were not able to either confirm or disprove this geometric mismatch hypothesis using such analytical approaches. Certainly, our CG simulations offer greater detail and resolution capable of capturing more subtle effects.

3.3. Backmapped, Atomistic Simulations. The CG model successfully demonstrated its ability to sample a wide variety of independent conformations and stabilize realistic ones, according to experimental data. To further validate the ensemble of thermally accessed conformations and obtain a refined picture of the stable folds, we ran backmapped, atomistic simulations of several low free-energy cluster centers presented above.

3.3.1. Hexamer (hex-at). Atomistic simulations of the barrel-forming hex-CG cluster #1 (Figure 2a and b) remain stable as shown by (*i*) the homogeneity of the protein chains' secondary structure map (Figure 5; note that the β-sheet regions stay

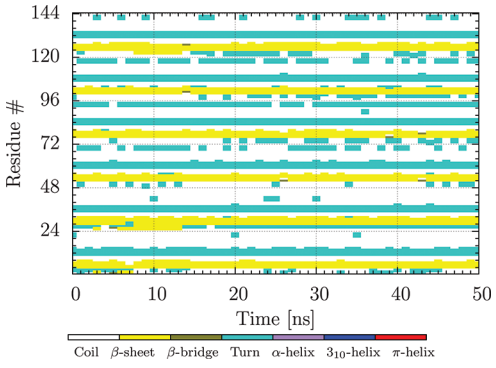

Figure 5. Secondary structure map (as assigned using stride[33]) of hex-at. β-sheet regions are well conserved throughout the simulation. Amino acids of the six protein chains have been stacked along the vertical axis. Restrained residues are #10–24, 34–48, 58–72, 82–96, 106–120, 130–144 (see Table 1).

present throughout the simulation) and (*ii*) a low C_α-RMSD ($\simeq$ 2 Å) from the initial conformation (Figure 6a). This simulation shows that the most stable cluster sampled in the hex-CG simulation stays stable when backmapped to an atomistic resolution. Overall, both CG and atomistic simulations are in agreement with SS-CCMV's X-ray structure.[4]

3.3.2. Pentamer (pent-at). In a similar manner, pent-CG clusters #1–4 (see Table 2, Figure 3, and Figure 4) were backmapped and simulated atomistically ("pent-at"). The secondary structure maps are shown in Figure 7 for clusters

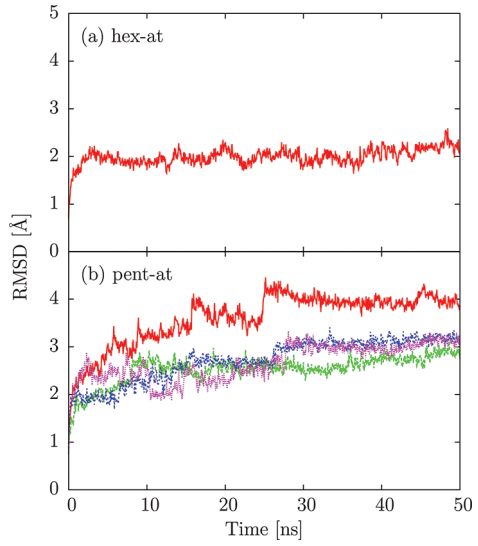

Figure 6. C_α-RMSD of (a) hex-at and (b) pent-at from their initial (backmapped) conformations: (a) hex-CG and (b) pent-CG#{1–4}, respectively. pent-CG#{1–4} correspond to four of the thermally accessible clusters of pent-CG (see Figure 3 for the relative free energies of these clusters).

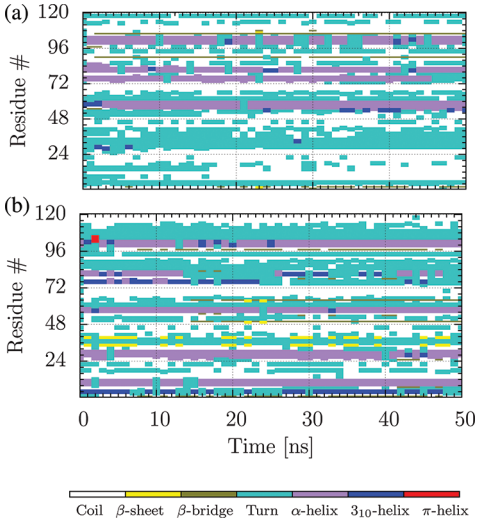

Figure 7. Secondary structure map (as assigned using stride[33]) of the three pent-at systems. pent-CG's clusters (a) #1 and (b) #3 were backmapped to provide initial conformations. Amino acids of the five protein chains have been stacked along the vertical axis. Restrained residues are #16–24, 40–48, 64–72, 88–96, 112–120 (see Table 1).

#1 and #3. While cluster #1 shows a modest variation in secondary structure over the course of the simulation from the initial conformation—indicating that the conformation is effectively locked in, cluster #3 changes a bit more. The C_α-RMSD of clusters #1–4, displayed in Figure 6, remains between $\simeq$2.5 and 4 Å. The moderate stability of each cluster, as established from free-energy calculations of the CG simulations (Figure 3c), suggests that the slow progress of the conformational change is not due to the presence of any stable free-energy minimum but rather due to slow kinetics, which prevent these atomistic simulations from sampling many independent configurations. We note that the absence of strong rearrangements during the course of these simulations suggests that the initial conformations (i.e., the pent-CG clusters) were

not widely out-of-equilibrium regarding the atomistic force field.

Overall, these results show that the atomistic time scales used here are largely insufficient to sample the system's conformational space. However, they permit the refinement of a nearly equilibrated initial conformation, as shown here from backmapping low free-energy CG cluster centers. Together, the secondary structure and RMSD analysis of the backmapped atomistic simulations show that these CG free-energy minima agree with what the atomistic force field considers "stable." This observation further validates and strengthens the usefulness of such a combined CG/atomistic approach in the context of flexible or unstructured protein systems.

4. CONCLUSIONS

The structural information inferred experimentally about SS-CCMV motivated our interest in understanding conformational aspects of parts of the viral capsid proteins: why does a 6-fold (i.e., hexameric) interface stabilizes a β-barrel, while the X-ray data[4] could not resolve any structure at the pentameric interface? Here, a combined CG/atomistic simulation approach was taken to thoroughly sample each interface and understand the thermodynamic implications of their geometry.

CG simulations of the hex system yielded a stable β-barrel conformation, compared to the other sampled clusters (Figure 3). A backmapped atomistic simulation of this CG cluster remained stable: the β-sheet regions were present throughout the 50 ns-long simulation (Figure 5).

CG simulations of the glycine-rich hexamer mutant (hex-gly-CG) reported both a large number of stable (β-rich) conformations and the absence of any β-barrel formation. This lack of stability was reported earlier by Willit et al.,[2] where dynamic light scattering experiments showed that the glycine-rich virion mutant disassembled at a temperature 10 °C below the native hexamer. They also pointed out that the presence of a β-barrel hexamer is not a necessary condition for virion self-assembly (although it does affect its kinetics).

CG simulations of the pentameric system (pent-CG) yielded a variety of stable helix-rich clusters, but no β-barrel. Subsequent backmapped-atomistic simulations of the three most stable CG clusters showed rather small changes, indicating that (*i*) the CG conformations were not widely out-of-equilibrium when run atomistically and (*ii*) the kinetics are slow compared to the simulation time scale. Furthermore, β-favoring CG simulations of pent provided a heterogeneous collection of β-rich conformations, but no barrel formation. This strongly suggests pent's inability to stabilize a β-barrel due to a geometric mismatch between the associated hydrogen-bond pattern and the chains' boundary conditions (i.e., how and where the relevant chains are positioned with respect to one another), rather than arising from the energetics of that conformation. This distinction between energetics and geometry in explaining the absence of a stable barrel at the 5-fold axis provides new insight into the stability of CCMV's capsid. This highlights how computational studies, in particular a mixed CG/atomistic approach combined with a determination of relative free energies of characteristic conformations, can complement experimental structure determination.

Speir et al. studied the stability of the SS-CCMV virus capsid using a protease-cleavage analysis.[4] Because the rate and locations of protein cleavage depend on the exposure of amino acids to the solvent, it allows the differential probing of protein regions. Their conclusions, partly based on the hydrogen-bonding network formed by the K42R mutation of SS-CCMV, suggest two sets of dynamic N-termini: strongly stabilized hexamer geometries and weakly stabilized pentamer geometries. Their results thus indicate the pentamer's higher propensity to access conformations that are more readily exposed to the solvent. We build on these experimental results to propose that pentamer geometries show not only thermally accessible solvent-exposing conformations (Figure 4) but also an important structural variability. While experimental techniques, such as X-ray or protease cleavage analysis, can tell whether a protein region is stable, the computational approach provided here gives insight into the *heterogeneity* of competing conformations. In this sense, the weak stability of the 5-fold interface is due to a large structural variability of thermally accessible conformations, rather than one or two competing structures.

The success of such a combined CG/atomistic study relies on a number of critical points that must be met: (*i*) long simulation times for accurate statistics, (*ii*) a CG model capable of secondary structure formation without explicit bias, (*iii*) a realistic α/β balance, (*iv*) reliable backbone dihedral motion to permit significant secondary structure variability, and (*v*) enough experimental information to constrain the boundary of the region of interest. It is because of all these criteria that the present study is successful in both reproducing expected results and bringing new insight into the stability of CCMV's capsid.

This combination of investigating the conformational free energy landscape on the CG simulation level and then locally exploring the minima with atomistic simulations will become more important in the future, for example also for the investigation of large protein complexes where protein/protein interfaces with structurally not well characterized regions may play a role for assembly and stability. The present results have shown that a computational study based solely on atomistic simulations would have been problematic in thoroughly sampling the relevant conformations. Here, we used CG simulations to determine and quantify the stability of the most stable conformations. Some of them were then backmapped and simulated atomistically to check whether strong structural rearrangements occurred, alleviating atomistic simulations' inherent sampling issues for the present systems. Moreover, the ability to backmap CG conformations to an atomistic resolution supplies material for direct structural comparison with several experimental techniques, such as NMR or cryo-electron microscopy.

AUTHOR INFORMATION

Corresponding Author

*E-mail: bereau@alumni.cmu.edu.

Present Address

[§]Department of Chemistry, University of Basel, Klingelbergstrasse 80, CH-4056 Basel, Switzerland

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We would like to thank Ananya Debnath and Olga Bezkorovaynaya for critically reading the manuscript. We are grateful to the Volkswagen Stiftung for supporting our project within the framework of the program "New Conceptual Approaches to Modeling and Simulation of Complex Systems".

C.P. also gratefully acknowledges financial support from the German Science Foundation within the Emmy Noether Program. Two of us (T.B. and M.D.) thank the MPI for Polymer Research, Mainz, for inviting us several times within the Theory group.

DEDICATION

We would like to dedicate this article to Wilfred van Gunsteren within this festschrift to honor his 65th birthday. Simulation algorithms, force fields, free energy calculations, analysis of conformational equilibria—the methods we all use in biomolecular simulation today—have seen essential contributions from Wilfred and his group. He has greatly advanced the entire field, bridging the disciplines from physics to chemistry to biology. Personally, C.P. is deeply indebted to Wilfred for giving her the chance to learn the trade of biomolecular simulation and, beyond that, also for being a role model and shaping her views on how to act responsibly as a scientist, colleague, and supervisor—and last but not least for spreading his apparently endless and inspiring optimism.

REFERENCES

(1) Zhao, X.; Fox, J. M.; Olson, N. H.; Baker, T. S.; Young, M. J. *Virology* **1995**, *207*, 486−494.
(2) Willits, D.; Zhao, X.; Olson, N.; Baker, T. S.; A. Zlotnick, J. E. J.; Douglas, T.; Young, M. J. *Virology* **2003**, *306*, 280−288.
(3) Speir, J. A.; Munshi, S.; Wang, G.; Baker, T. S.; Johnson, J. E. *Structure* **1995**, *3*, 63−78.
(4) Speir, J. A.; Bothner, B.; Qu, C.; Willits, D. A.; Young, M. J.; Johnson, J. E. *J. Virol.* **2006**, *80*, 3582−3591.
(5) Zandi, R.; Reguera, D.; Bruinsma, R. F.; Gelbart, W. M.; Rudnick, J. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 15556−15560.
(6) Praprotnik, M.; Delle Site, L.; Kremer, K. *Annu. Rev. Phys. Chem.* **2008**, *59*, 545−571.
(7) Peter, C.; Kremer, K. *Soft Matter* **2009**, *5*, 4357−4366.
(8) Terakawa, T.; Takada, S. *Biophys. J.* **2011**, *101*, 1450−1458.
(9) Kamerlin, S. C.; Vicatos, S.; Dryga, A.; Warshel, A. *Annu. Rev. Phys. Chem.* **2011**, *62*, 41−64.
(10) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33−38.
(11) Lucas, R. W.; Larson, S. B.; McPherson, A. *J. Mol. Biol.* **2002**, *317*, 95−108.
(12) Stockley, P. G.; Twarock, R. *Emerging Topics in Physical Virology*; World Scientific Publishing Co.: Hackensack, NJ, 2010.
(13) Sali, A.; Blundell, T. L. *J. Mol. Biol.* **1993**, *234*, 779−815.
(14) Bereau, T.; Deserno, M. *J. Chem. Phys.* **2009**, *130*, 235106.
(15) Go, N. *Annu. Rev. Biophys. Bioeng.* **1983**, *12*, 183−210.
(16) Bereau, T.; Bachmann, M.; Deserno, M. *J. Am. Chem. Soc.* **2010**, *132*, 13129−13131.
(17) Bereau, T.; Deserno, M.; Bachmann, M. *Biophys. J.* **2011**, *100*, 2764−2772.
(18) Pautsch, A.; Schulz, G. E. *Nat. Struct. Mol. Biol.* **1998**, *5*, 1013−1017.
(19) Limbach, H.-J.; Arnold, A.; Mann, B. A.; Holm, C. *Comput. Phys. Commun.* **2006**, *174*, 704−727.
(20) Keller, B.; Daura, X.; van Gunsteren, W. F. *J. Chem. Phys.* **2010**, *132*, 074110.
(21) Ferrenberg, A. M.; Swendsen, R. H. *Phys. Rev. Lett.* **1988**, *61*, 2635−2638.
(22) Ferrenberg, A. M.; Swendsen, R. H. *Phys. Rev. Lett.* **1989**, *63*, 1195−1198.
(23) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 1011−1021.
(24) Bereau, T.; Swendsen, R. H. *J. Comput. Phys.* **2009**, *228*, 6119−6129.
(25) Chernick, M. R. *Bootstrap Methods: A Guide for Practitioners and Researchers*, 2nd ed.; John Wiley & Sons: New York, 2008.
(26) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(27) Schmid, N.; Eichenberger, A.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A.; van Gunsteren, W. F. *Eur. Biophys. J.* **2011**, *40*, 843−856.
(28) Berendsen, H. J.; Postma, J. P.; van Gunsteren, W. F.; Hermans, J. *Intermolecular Forces*; D. Reidel Publishing Company: Dordrecht, The Netherlands, 1981; pp 331−342.
(29) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.
(30) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.
(31) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, L. G.; anf Pedersen, H. *J. Chem. Phys.* **1995**, *103*, 8577.
(32) Hess, B. *J. Chem. Theory Comput.* **2008**, *4*, 116−122.
(33) Frishman, D.; Argos, P. *Proteins: Struct. Funct. Genet.* **1995**, *23*, 566−579.
(34) Berg, J. M.; Tymoczko, J. L.; Stryer, L. *Biochemistry*, 6th ed.; W. H. Freeman & Co.: New York, 2006.
(35) Lasters, I.; Wodak, S. J.; Alard, P.; van Cutsem, E. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 3338−3342.
(36) Murzin, A. G.; Lesk, A. M.; Chothia, C. *J. Mol. Biol.* **1994**, *236*, 1369−1381.