Tabulation as a high-resolution alternative to coarse-graining protein interactions: Initial application to virus capsid subunits

Justin Spiriti and Daniel M. Zuckerman^{a)}

Department of Computational and Systems Biology, University of Pittsburgh, 3501 Fifth Ave., Pittsburgh, Pennsylvania 15260, USA

(Received 4 September 2015; accepted 10 December 2015; published online 30 December 2015)

Traditional coarse-graining based on a reduced number of interaction sites often entails a significant sacrifice of chemical accuracy. As an alternative, we present a method for simulating large systems composed of interacting macromolecules using an energy tabulation strategy previously devised for small rigid molecules or molecular fragments [S. Lettieri and D. M. Zuckerman, *J. Comput. Chem.* **33**, 268-275 (2012); J. Spiriti and D. M. Zuckerman, *J. Chem. Theory Comput.* **10**, 5161-5177 (2014)]. We treat proteins as rigid and construct distance and orientation-dependent tables of the interaction energy between them. Arbitrarily detailed interactions may be incorporated into the tables, but as a proof-of-principle, we tabulate a simple α -carbon G $\ddot{o}$ -like model for interactions between dimeric subunits of the hepatitis B viral capsid. This model is significantly more structurally realistic than previous models used in capsid assembly studies. We are able to increase the speed of Monte Carlo simulations by a factor of up to 6700 compared to simulations without tables, with only minimal further loss in accuracy. To obtain further enhancement of sampling, we combine tabulation with the weighted ensemble (WE) method, in which multiple parallel simulations are occasionally replicated or pruned in order to sample targeted regions of a reaction coordinate space. In the initial study reported here, WE is able to yield pathways of the final $\sim 25\%$ of the assembly process. © 2015 AIP Publishing LLC. [<http://dx.doi.org/10.1063/1.4938479>]

INTRODUCTION

A variety of coarse-grained (CG) force fields have been developed for biomolecular systems^{1–31} in order to simulate larger systems for longer than would be possible with molecular dynamics (MD) simulation. Because of the fundamental importance of protein–protein interactions to biology, several methods have been developed specifically for CG simulations of multiprotein systems. Levy *et al.* studied small protein oligomers by means of a G $\ddot{o}$ -like model based on native interactions similar to the one employed in the present work and demonstrated that this model could reproduce important features of the binding mechanism determined from laboratory experiments.^{32,33} Elcock and co-workers constructed a molecular model of the cytoplasm consisting of some ~ 50 different types of macromolecules in order to study the effect of macromolecular crowding; the model included collective charges for each macromolecule derived from the electrostatic potential calculated using Poisson-Boltzmann methods, combined with a Lennard-Jones interaction potential.³⁴ The same group is constructing a CG force field for proteins based on atomistic simulations of amino acid interactions.^{35,36} Several other groups have also developed CG force fields expressly for simulating protein–protein interactions or macromolecular crowding, either basing them on all-atom force fields³⁷ or calibrating them to reproduce the structures of specific protein dimers.^{38,39} The specific

applications of CG force fields to protein–protein interactions have been reviewed in recent years.^{21,40}

It is well-appreciated that CG models typically offer two key advantages and one disadvantage.^{16,20,41–43} The benefits are a reduction in computing time—per energy call or time step—during a simulation, as well as smoothing of the energy landscape—thus reducing the number of time steps required for a given level of sampling.⁴¹ On the other hand, CG models typically omit fine-grained details which may be required for chemical accuracy.

We have been developing an energy-tabulation strategy^{44,45} that may be seen as related to coarse graining, but which has somewhat distinct advantages and disadvantages. Beneficially, tabulation can reduce the cost of energy calculation, it can yield smoothed energy landscapes, and it may sacrifice chemical accuracy only minimally—with the latter two depending on the system and implementation. However, tabulation typically requires eliminating or compromising some molecular flexibility, and so must be employed in a judicious fashion.

The tabulation approach entails dividing a system into fragments which will be assumed rigid, and the interaction energies between every possible relative displacement and orientation of these fragments are computed—using an arbitrary force field—and stored in tables. During subsequent simulations, the interaction energies can then be looked up from these tables, rather than being recalculated. The ability of tables to speed the energy calculation depends on the size of the fragments. While the method was able to achieve a 20-fold speedup in simulations of benzene using

^{a)} Author to whom correspondence should be addressed. Electronic mail: ddmmzz@pitt.edu

an all-atom force field,⁴⁴ it was not possible to achieve a significant speedup when working with proteins in a united-atom force field, because amino acids had to be divided into relatively small fragments containing an average of 4 atoms each.⁴⁵ Nevertheless, in both the previous studies, tabulated simulations retained a significant degree of atomistic accuracy.

Here, the tabulation approach will be employed to study viral capsids which, because of their biomedical importance, previously have been simulated using a wide range of techniques. Some of the largest all-atom MD simulations performed to date have been of viral capsid assembly. For example, the Schulten group performed all-atom simulations of the satellite tobacco mosaic virus for 50 ns, finding that the presence of RNA is required for stability.⁴⁶ A subsequent comparison of the all-atom simulation to CG simulation of the same virus on the μ s time scale, indicated that the two simulations exhibit similar behavior.⁴⁷ Later, the same group took a similar approach to simulate the complete, much larger HIV-1 capsid as well as smaller portions and a tubular assembly formed from capsid subunits,⁴⁸ demonstrating the role of intersubunit interfaces in ensuring the proper angles between subunits needed for capsid stability.

Our interest here extends beyond the short-time scale thermal fluctuations accessible via MD simulation and instead focuses on the much slower capsid assembly process, which also has received significant theoretical attention.^{49–57} Although a great deal has been learned from experimental biophysical studies of the assembly of many viruses, it is difficult to detect intermediates and precisely characterize the assembly mechanism experimentally.⁵⁸ (However, some useful information on the sizes of intermediates has been obtained by ion mobility-mass spectrometry.⁵⁹) Simulations have provided important information about assembly mechanisms, such as relationships between the strength of intersubunit interactions, the T-number, and the type of assembly process.^{51,54} Because of the large size of viral capsids (on the order of 100 Å) and long time scales associated with assembly (on the order of 100 s⁶⁰), these simulations have used heavily coarse-grained representations, in which, for example, a viral subunit is represented by a “patchy sphere” or a small collection of no more than 20 spherical particles. Even with such severe coarse graining, however, these simulations can take a considerable amount of computer time. For example, simulations of icosahedral capsid assembly by discrete molecular dynamics required approximately 30 central-processing-unit (CPU) days to reach a full equilibrium.⁵¹ Consequently, it may be quite challenging to improve the structural detail of the simulations unless new computational approaches are employed, motivating our use of tabulation.

In this first application to a multi-protein system, we study the hepatitis B viral capsid using the tabulation strategy. The system was chosen because it is known that the capsid protein can assemble *in vitro* without any other viral components, and because experimental thermodynamic^{61,62} and kinetic^{60,63} data regarding the assembly process as well as a crystallographic structure of the complete capsid⁶⁴ are available. Also, all of

the subunits are identical in sequence and approximately the same in structure (the α -carbon root-mean-square deviation (RMSD) between subunits ranges from 0.8 to 1.2 Å), simplifying the model. By defining subunit dimers as the rigid fragments whose relative displacement and orientation form the basis for the tables, we are able to reduce the computational cost of interaction-energy calculation by at least three orders of magnitude compared to a simulation in which energies were directly calculated without tables. This allowed us to use a Gō-like model^{3,4,33} in which each α -carbon in the subunit is represented. Although not atomistic, this model is significantly more spatially realistic than previous models used for capsid assembly,^{49,51,52,54,55} since each dimer is represented by a total of 284 particles, compared to at most 20 or so instead of the heavily coarse-grained models that have typically been used. We emphasize that more detailed models can be tabulated.

Even a fast and accurate model can still be limited by intrinsic, model-specific time scale limitations for a slow process like capsid assembly—necessitating the use of additional sampling procedures. Therefore, in addition to the tables, we used the weighted ensemble (WE) approach,^{65–67} as implemented in the WESTPA software.⁶⁸ In WE simulation, configuration space is divided into bins based on progress coordinates (here the degree of assembly and proximity to the assembling partial capsid), with the goal of sampling trajectories in all bins. Multiple trajectories are run in parallel and periodically replicated or pruned—but never biased—in order to maintain a target number of simulations in each bin; in this way, WE allots computing resources to favor assembly. Using this method, we have been able to obtain trajectories simulating the final stages of assembly of the hepatitis B viral capsid.

The combination of tabulation and WE simulation has yielded encouraging results in this initial application to the multi-protein scale. In particular, we have been able to simulate the final stages of capsid assembly. We have obtained trajectories progressing from 110 to 120 dimers (a complete capsid) and also from a $3/4$ capsid with 90 dimers to a nearly complete 117.

The remainder of the paper is divided into several sections. The first describes the underlying energy function and its tabulation, as well as the nature of the progress coordinates and binning for WE. Following this, the results from test simulations conducted to study computational speedup as a function of the number of Gō model sites and capsid stability as a function of table resolution and size are presented. The results from the WE simulation are then discussed, with particular attention to the abovementioned assembly pathways. Finally, we discuss the limitations of this first attempt and future prospects for obtaining a spatially realistic simulation of the full assembly of hepatitis B capsid and similar systems.

METHODOLOGY

Three basic methods were employed: (i) parameterization of a Gō-like model for capsid protein–protein interactions, (ii)

tabulation of a force field into look-up tables of a few gigabytes (GB) in size for the assumed-rigid capsid protein dimers of two types, and (iii) weighted ensemble simulation of the tabulated force field using Metropolis Monte Carlo (MC).

Energy function and tabulation

Although almost any model can be incorporated into tables, in this first test, we chose to use a Gō-like model for protein–protein interactions similar to that used by Levy *et al.*^{32,33} but constructed to ensure stability of the complete capsid as represented in the experimental structure. We constructed interaction tables similar to those previously used for proteins⁴⁵ based on this Gō-like potential. These tables were then used for Monte Carlo simulations to run tests. In addition, weighted ensemble Monte Carlo simulations were performed, starting from a partially formed capsid surrounded by a bath of free dimers. During the simulations, dimers from the bath added to the capsid, revealing a possible pathway for assembly. The parameters for all simulations are as listed in Table I, with exceptions as indicated below in the descriptions of individual simulations.

Capsid structure preparation

The coordinates in the original Protein Data Bank (PDB) file (code 1QGT) were expanded to form a complete structure of the capsid, using the transformations listed in the PDB file. Only the first 142 residues of the subunit were used, since coordinates for only these residues were available. The subunits are joined together in pairs by disulfide bridges to form two dimers per asymmetric unit, for a total of 120 dimers. In the PDB file, one dimer is labeled with chain codes A and B (an AB dimer) and the other with chain codes C and D (a CD dimer). The dimers have similar but slightly different conformations. The AB dimers form the vertices with fivefold symmetry, whereas the CD dimers form vertices with threefold symmetry; at sites with approximate sixfold symmetry, two AB dimers and four CD dimers come together (Fig. 1(a)). Each of these dimers was treated as a rigid fragment, containing a total of 284 Gō model sites (α -carbons) per dimer as shown in Fig. 1(c). The AB and CD dimers were treated as distinct types of fragments in order to maintain the ability to form 5-fold or 3-fold symmetry as needed.

Energy function

In this initial study, we chose to use a Gō-like model^{3,4} to represent the interactions between rigid dimers that result in their assembly. This potential is based on the structure of the fully assembled capsid, but does not include separate, explicit terms representing physical forces such as the hydrophobic or electrostatic forces known to be important in hepatitis B capsid assembly.⁶⁹ Instead, it is an effective potential that assumes nearby native sites that are mutually attractive in a general way. Nevertheless, a Gō model ensures that the fully assembled capsid represents a stable potential-energy minimum, and it is computationally inexpensive, facilitating the construction of the interaction energy tables described below. Similar

potentials are commonly used to study the dynamics of protein folding and conformational change.^{32,66,70–73}

The interaction energy between two rigid dimers separated by a displacement with spherical coordinates (r, θ, φ) and relative orientation with Euler angles $(\varphi', \theta', \psi')$ has the following form:^{3,4,32,33}

$$U(r, \theta, \phi, \phi', \theta', \psi') = \sum_{ij} \left\{ \begin{array}{l} \varepsilon \left[5 \left(\frac{r_{ij}^0}{r_{ij}} \right)^{12} - 6 \left(\frac{r_{ij}^0}{r_{ij}} \right)^{10} \right] \text{ (native contacts)} \\ 5\varepsilon \left(\frac{2r_{\text{HC}}^0}{r_{ij}} \right)^{12} \text{ (nonnative contacts)} \end{array} \right. \quad (1)$$

where the sum is taken over all inter-fragment pairs of α -carbons, and to speed evaluation, terms in which $(r_{ij}/r_{ij}^0)^2 > 2$ or $(r_{ij}/2r_{\text{HC}})^2 > 2$ were omitted. Here, r_{ij} is the distance between atoms i and j , and the native distance r_{ij}^0 is determined as the minimum distance between atoms i and j over all possible pairs of fragments of those types in the crystal structure. A contact was considered to be native if r_{ij}^0 was less than the cutoff 10 Å, and r_{HC} is a hard-core radius which was set at 1.7 Å. The remaining parameters are described and values are given in Table I. (For some timing tests, Gō models were constructed which used only some of the α -carbons as sites, together with a larger hard-core radius, as described below.)

We performed test simulations varying the well depth ε (regulating the overall strength of interactions between dimers) and the distance cutoff for determining native interactions in test simulations. We found that interactions that are too weak and/or distance cutoffs that are too short resulted in disassembly of the capsid, while interactions that are too strong and/or distance cutoffs that are too long led to an unphysical attachment of “polymers” of dimers to the outside of the partially formed capsid. At the end, a well depth of 0.3 kcal/mol and a native distance cutoff of 10 Å were chosen which appeared to promote assembly of correctly formed capsids as effectively as possible.

Tables

In order to speed Monte Carlo simulations conducted using the potential of Eq. (1), interaction energy tables were constructed in terms of the relative displacement and orientation of two dimers, expressed in terms of spherical coordinates and Euler angles. The tables were constructed in a similar manner to those described previously for proteins; in particular, the radial resolution was on an exponential scale.⁴⁵ The relative displacements used were between the centers of mass of the two fragments; the orientations of each fragment were measured relative to the principal axes of the moment of inertia of each fragment. Test simulations of a complete capsid were conducted with tables of several different resolutions in order to study the dependence of capsid stability on table resolution. These simulations used a capsid in periodic boundary conditions with a cubic box of edge length

FIG. 1. (a) α -carbon representation of a complete hepatitis B viral capsid, from PDB structure 1QGT. ((b) and (c)) Comparison of different representations of a single AB dimer. (b) All-atom representation. (c) Representation in Gō-like model, in which each α -carbon is a separate particle.

2711 Å and were run for 1×10^9 trial moves. Based on these simulations, a high angular resolution was chosen for the tables used for the subsequent simulations. To further test the stability of the partially formed capsid, another test simulation was performed, starting from a partially formed capsid of 110 dimers surrounded by a bath (the starting structure for simulation 1, described below). This simulation was performed for approximately 11×10^9 trial moves. Because the two types of fragments were used, three tables were needed for these simulations, with a total size of approximately 13.5 GB.

Table resolution and speedup studies

To determine the relationship between the number of Gō particles per fragment and the extent of computational speedup, several additional sets of tables were constructed. Tables were generated with only every 2nd, 5th, 10th, or 20th α -carbon as Gō model sites (for a total of 142, 58, 30, or 16 sites per dimer, respectively). Short timing tests of 1×10^9 trial moves were run both using these tables and using the same models without tables, in order to compare the speed of the simulation in each case.

TABLE I. Summary of simulation parameters. Parameters in boldface are different between parameter sets 1 and 2.

Symbol	Description	Value (parameter set 1)	Value (parameter set 2)
ϵ	Gō model well depth	0.2 kcal/mol	0.3 kcal/mol
r_{HC}	Hard core radius	1.7 Å	1.7 Å
	Cutoff for defining native contacts in Gō model	10 Å	10 Å
	Interfragment energy cutoff	100 Å	100 Å
	Radial resolution of tables	20-100 Å	20-100 Å
	Radial table resolution	0.5-2.5 Å (exponential scale ⁴⁵)	0.5-2.5 Å (exponential scale ⁴⁵)
	Spherical angular table resolution (θ , ϕ)	5°	5°
	Euler angle resolution (ϕ' , θ' , ψ')	15°	15°
	Maximum translational MC trial move size	0.5 Å	0.5 Å
	Maximum rotational MC trial move size	2.0 Å	2.0 Å
	Distribution of MC moves	50% translational, 50% rotational	50% translational, 50% rotational
	Number of dimers in simulation	300	300
	Dimer concentration (edge length of cubic box)	250 μM (1258.39 Å)	250 μM (1258.39 Å)
	MC simulation temperature	300 K	300 K
	Initial number of dimers in central capsid (surrounding bath)	110 (190)	90 (210)
τ	Number of MC trial moves per WE iteration	10^6	10^7
R	Distance cutoff for partially formed capsid identification algorithm	80 Å	80 Å
α	Orientational cutoff for partially formed capsid identification algorithm	30°	30°
	Number of trajectories per bin	16 for close bins, 4 for others (see Fig. 2)	16 for close bins, 4 for others (see Fig. 2)
	Frequency of merging of bins into catch-all bin	Every 50 WE iterations	Every 50 WE iterations
	Rule for bin merging (see Figure 2)	Leave 1 bin behind most-assembled segment	Leave 2 bins behind most-assembled segment

FIG. 2. (a) Example initial system configuration for simulation 1. (b) Close up view, showing nearly formed capsid with 110 fragments at center.

Weighted ensemble simulations

To simulate capsid assembly, weighted ensemble simulations^{65–68} were conducted using a binning strategy that encouraged capsid assembly. In the weighted ensemble method as used here, configurational space is divided into bins defined via one or more progress coordinates. Multiple simulations are assigned initially equal weights and run in parallel; occasionally, simulations are “split” (replicated to preserve local weight) and merged (pruned in a weight-preserving way) as needed in order to maintain a target number of simulations in each bin. In this way, trajectories diffuse in the space defined by the progress coordinates. Since the progress coordinates we use here measure the extent of capsid assembly, WE simulation can yield trajectories that represent possible assembly pathways. WE simulations were conducted using the WESTPA software,⁶⁸ version 1.0.0 beta.

In this proof-of-principle study, a partially formed capsid was incorporated into the starting structure for each simulation, so that it would not be necessary to overcome any barrier associated with the nucleation of capsids. This partially formed capsid was surrounded by a bath of additional dimers in a simple cubic lattice with random initial orientations. (In all cases, an equal number of AB and CD dimers were present in the simulation.) The central capsid initially contained either 90 or 110 dimers (out of 120 in a complete capsid). The nearly complete capsid with 110 dimers was obtained by removing 10 dimers around a fivefold symmetry axis from a complete capsid, as shown in Figure 2. The nearly complete capsid with 90 dimers was obtained by removing three groups of 10 dimers around a threefold symmetry axis. From these structures, Monte Carlo simulations were performed using moves that included random translations and rotations on individual dimers. Depending on whether the simulations started with 90 or 110 dimers, different sets of parameters for Monte Carlo and WE were used; three replicate simulations were performed with each parameter set. The parameter sets are also listed in Table I, and the length of each simulation is shown in Table II.

In order to define progress coordinates for weighted ensemble that promote capsid assembly, it is first necessary to be able to identify partially formed capsids automatically. An algorithm was devised for this purpose, which is described in detail in the Appendix. Briefly, this algorithm alternates between RMSD fitting⁷⁴ of fragment centers in frames from the simulation to corresponding fragments in a fully

formed capsid and updating the correspondence by assigning fragments from the capsids to the closest fragments in the fitted capsid. A distance and an angle cutoff used in the latter reassignment process determine how closely partially formed capsids must resemble the crystal structure of a fully formed capsid in order to be recognized during this simulation.

Two progress coordinates were used in all WE simulations. The first coordinate was the number of dimers in the largest partially formed capsid (which invariably was the original nearly complete capsid placed in the initial structure). The second coordinate was the minimum distance from the center of any fragment not already in the central capsid to an empty spot on the template as aligned to the central capsid. (This progress coordinate was included in order to accelerate the movement of dimers toward the central, incomplete capsid.) Figure 3 shows the manner in which the space of these two progress coordinates was divided into bins.

Several additional strategies were employed to further limit the amount of computer time needed for the WE simulations. The target number of trajectories per bin was set non-uniformly among the bins so that those bins in which the minimum distance coordinate was greater than 200 Å had fewer trajectories (a target of 4 trajectories per bin) compared to those in which the minimum distance fragment was closer to the capsid (16 trajectories per bin). In addition, at intervals of every 50 iterations during the course of the simulation, bins in the direction of the first reaction coordinate (size of the partially formed capsid) were absorbed into the catch-all bin. Since the WE algorithm automatically maintained a target of 4 trajectories in this bin, most of the trajectories were

TABLE II. Length of WE simulations conducted using parameter sets listed in Table I, and number of dimers in the central capsid at the beginning and end of each simulation, as determined by the RMSD-based algorithm.

Parameter set	Simulation	Number of WE iterations	Initial number of dimers in capsid	Final number of dimers in capsid
1	1	1000	110	120
1	2	500	110	120
1	3	600	110	120
2	4	1000	90	117
2	5	1000	90	113
2	6	550	90	117

FIG. 3. Schematic diagram showing the binning strategy for the WE simulations, as well as the manner in which bins were merged into the catch-all bin during the simulation.

eliminated at each merger, leaving behind only those that had made the most progress toward assembly. These two strategies concentrated the computational effort on those trajectories that had made the most progress toward complete assembly and that were most likely to add an additional dimer to the central capsid.

Y132A mutation and nucleation simulations

To investigate the degree to which an alpha-carbon Gō model was sufficient to represent some aspects of residue-level chemical specificity, we examined the Y132A mutation of the capsid subunit, which has been shown experimentally to inhibit capsid formation.⁶³ While the overall shape of Y132A mutant subunit dimers is similar to that of wild-type dimers, small but significant changes in the conformation cause the mutant dimers to assemble into trimers of dimers rather than complete capsids.⁶³ We constructed a Gō model for the mutant dimer similar to that for the wild type dimer. The wild-type native geometry and almost all of the wild-type native

interaction distances were used, while the native interactions involving residue 132 were removed. Although this does not take into account the minor but important changes in dimer conformation caused by the mutation, it facilitates comparison with our wild-type model.

To investigate the effect of the Y132A mutation on capsid nucleation, brute force simulations of both wild-type and mutant dimers were performed, starting from a lattice arrangement of isolated dimers with a concentration of 250 μ M and no embedded partially formed capsid. Five replicate simulations of each type of dimer were conducted for a total of 10×10^9 trial moves each. They were analyzed using the same RMSD-based algorithm as was used for the WE simulations. However, instead of using a complete capsid as a template, collections of either five dimers arranged in fivefold symmetry, or three dimers arranged in threefold symmetry, drawn from the structure of the capsid, were used as templates. A distance cutoff of 10 Å and an angular cutoff of 30° were used. Based on this analysis, nuclei were defined as containing either three out of five dimers in fivefold symmetry (a pentameric nucleus) or all three dimers in threefold symmetry (a trimeric nucleus).

RESULTS

In overview, we first examine the accuracy and stability of the tabulated simulations, and then study the relationship between computational speedup and the number of Gō model particles. We then describe the assembly simulations obtained using WE, focusing on the order in which dimers attach to the partially formed capsid.

Table resolution and accuracy

Figure 4 shows the effect of tabulation on the interaction energy between two dimers (containing chains A and B) as a function of their relative displacement. The energy isosurfaces of the tabulated potential are broadly similar to the non-tabulated potential, although the discrete, stepwise

FIG. 4. Isosurfaces of interaction energy between two AB dimers having the same orientation. (a) Illustration of two coarse-grained representations of the dimer, showing relative position and orientation in the isosurface plots. (b) Interaction energy between wild-type dimers as calculated directly without tables and (c) as represented in high-resolution table (radial resolution 0.5 Å, angular (θ , ϕ) resolution 5°, and orientational (ϕ' , θ' , ψ') resolution 15°). (d) Interaction energy between Y132A mutant dimers calculated without tables, as in (b), showing increased repulsion near sharp corners. In (b)–(d), the blue contour represents an interaction energy of +1 kcal/mol; the red contour represents an interaction energy of −1 kcal/mol. Arrows in (b) and (d) indicate the position of residue 132 in each subunit within the dimer.

FIG. 5. Average fragment center RMSD as a function of table size for capsid test simulations with tables of various resolutions. Data point labels indicate table resolution, given in terms of radial resolution in Å, then angular and finally orientational resolution. (Inset) Time series of fragment center RMSD for the 0.5-5-15 resolution.

nature of the tabulation is also evident. The α -carbon-scale granularity of the non-tabulated potential is not fully discernible in the tabulated potential because the α -carbon separation is approximately of the same length as the tabulation discretization scale in the present implementation. The issue of resolution limitations in tabulation, which will depend both on the system and on hardware, is pursued further in the Discussion section.

Figure 5 summarizes stability results of Monte Carlo simulations starting from the *complete* capsid using tables of different resolutions (and differing amounts of computer memory). Although all simulations yielded relatively stable capsids, there is a definite correlation between the fragment center RMSD relative to the original structure and the table resolution. A higher table resolution corresponds to a smaller deviation of the capsid from the starting structure. The angular and orientational resolutions appear to be particularly important, since a simulation with 0.5 Å radial resolution, but 15° angular and 20° orientational resolutions produced a higher RMSD than a simulation with 1.0 Å radial resolution,

FIG. 6. Time series of the fragment center RMSD for the nearly complete capsid within a brute force simulation started from an initial structure with 110 dimers in the central capsid, surrounded by a bath.

FIG. 7. Energy calculation speedup factor as a function of the number of Gō model sites per fragment. The scaling exponent is 1.58 ± 0.08 .

but 10° angular and 15° orientational resolutions. This is because the higher angular and orientational resolutions for the tables better represent the angles at which the dimers must come together to form a capsid with icosahedral symmetry. In addition, the simulations are stable, as exemplified by the time series of fragment center RMSD shown in the inset.

Figure 6 shows the average fragment RMSD for the central capsid in the brute-force simulation started with a partial capsid of 110 dimers and a bath. This simulation was also stable, and over its course, one dimer added to the original 110 dimers over the course of approximately 11×10^9 trial moves. This suggests that a very substantial single trajectory would be required to obtain a complete capsid.

Energy calculation speedup

Figure 7 and Table III show a comparison of the computational speed with and without tables, as a function of the number of Gō model sites. In contrast to our previous experience with proteins,⁴⁵ we obtain an increase of energy calculation speed of at least a factor of 1000 when every α -carbon is used as a site without needing to alter the interaction potential via smoothing. This is primarily because the rigid fragments have a large number of sites, namely 284 Gō model particles, compared to the average of about 4 particles per rigid fragment in our fragment-based protein models.⁴⁵ When tables are not used, the number of interparticle distances that must be calculated in order to determine the interaction energy between

TABLE III. Computational speedup factor variation with number of interaction sites. Speeds are in trial moves per second.

Number of Gō model sites	Speed without tables	Speed with tables	Speedup factor
16	49 677.10	735.29	68
30	49 236.83	270.27	182
58	49 529.47	134.41	368
142	49 751.24	26.22	1897
284	49 751.24	7.19	6781

FIG. 8. Time course and order of dimer addition to the central capsid over the course of simulations 1-3, starting with 110 dimers. Each image of a capsid is color coded according to the WE iteration number when it became attached to the capsid in an example trajectory within one of the simulations (gray dimers were part of the initial structure).

two dimers is equal to the square of the number of particles in a fragment. However, because a secondary cutoff was used on the ratio of the interparticle distance to either the native or hard-core distance, not all of these interactions require a full calculation of the interaction energy. Consequently, it is expected that the computational speedup will grow somewhat slower than the square of the number of sites (aside from hardware-specific effects). A log-log linear regression gives a scaling exponent of 1.58 ± 0.08 . The outlook for using more realistic models is pursued in the Discussion.

Final events in viral capsid assembly

Both of the weighted ensemble simulation sets showed significant additional assembly from their respective starting

structures. Figure 8 shows the time course and order in which dimers were added to the central capsid in simulations using parameter set 1, in which the central capsid started with 110 dimers. In each case, at the end of the simulation, the RMSD-based algorithm identified 120 dimers as being part of the capsid. However, visual inspection shows that, in simulations 2 and 3, only 119 dimers were actually part of the capsid, with one dimer floating outside the capsid but close enough to the empty slot to be counted as part of the capsid by the algorithm. The ordering of the dimer addition also varied from one simulation to another. In simulations 1 and 2, dimer addition proceeded from the outside in, while in simulations 2 and 3, dimer addition proceeded generally from one side of the “hole” in the initial structure to the other, while leaving small gaps that are subsequently filled in.

FIG. 9. Time course and order of dimer addition to the central capsid over the course of simulations 4-6, starting with 90 dimers. Each image of a capsid is color coded according to the WE iteration number when it became attached to the capsid in an example trajectory within one of the simulations (gray dimers were part of the initial structure).

The simulations using parameter set 2, starting from 90 dimers, also reached nearly complete assembly in all three cases, with 113–117 dimers in the central capsid at the end of the simulation, as shown in Figure 9. In simulations 4 and 5, the remaining unfilled cavities are isolated from one another, so that the final configurations are not analogous to those with the same number of dimers in simulations starting from 110 dimers. The inability to fill in these defects appears to be connected to the lack of isolated dimers in the maturing bath, which not only has been depleted in absolute terms but also has become enriched in higher-order species. Simulation 6 shows a somewhat different final structure, in which, rather than closing up, the capsid develops two partially overlapping “levels” of dimers. In all three cases, the simulations generally assemble from the outside as expected, with the first dimers adding to the edge of the “hole” created in the initial structure and later dimers filling in the interior of that hole. However, the assembly is not uniform and the precise sequence of additions varies from one simulation to another.

The mechanism of dimer addition in all simulations consisted solely of single dimer addition to the existing partial capsid in these initial simulations. As expanded in the Discussion, this could be due to any of the several factors: the force field, the WE progress coordinates, or the starting configuration—including the relatively “immature” bath, which may not have had sufficient time to form higher-order structures. Although WE simulation is unbiased, the short overall length of individual trajectories may initially yield uncharacteristic trajectories, as we describe further in the Discussion section.

Y132A mutant potential and simulations

We also examined the Y132A mutation, for which an interaction energy isosurface is shown in Figure 4(d). The mutant potential is identical to the wild-type potential except for having repulsive regions on the sharp exterior corners where Y132 is located.

The number of clusters identified in both wild-type and mutant lattice simulations is shown as a function of the number of trial moves in Figure 10. The rate at which both pentameric and trimeric nuclei are formed during the simulation is substantially lower for the Y132A mutant dimer

than for the wild-type dimer. The reduction in the formation of pentameric nuclei is consistent with the experimental inability of the Y132A mutant to assemble into complete capsids. However, the reduction in the formation of trimeric nuclei is not as consistent with this experimental finding, since the Y132A mutant still forms trimeric assemblies of dimers. This could be due to the force field or due to the use of the wild-type conformation as a reference geometry for the mutant dimer. Nevertheless, these results demonstrate the ability of a residue-level approach to model the effects of individual mutations, something beyond the capabilities of lower resolution models.

DISCUSSION

This proof-of-principle study combining tabulation and WE simulation is an encouraging first step in developing an alternative to traditional site-based coarse-graining for multi-protein system. The tabulation scheme speeded energy calculation by a factor exceeding 10^3 , enabling a level of resolution not previously practical for capsid assembly, and WE simulations provided partial-assembly events which otherwise would not have occurred using the present energy function even with the tabulation speedup. The goal of the present study was to determine whether the tabulation strategy is promising for future, more realistic studies. At this point, comparison to experimental data based on Monte Carlo simulation of the over-simplified Gō-like energy function does not seem warranted, and indeed, a number of improvements to our initial protocol are warranted.

Although we have not yet simulated a complete assembly, and despite the energy model’s simplicity, the pathways obtained here offer a preview of the kind of insights that a higher resolution simulation of capsid assembly might provide. For example, the asymmetric, non-uniform way in which dimers are added in many of the simulations suggests the possibility that assembly intermediates may not share the symmetry of the final capsid structure. Extending the simulations to start from progressively fewer dimers in the central capsid should yield more complete pathways that yield even more information. Such pathways could be compared with similar pathways generated by other approaches, such as by using the Gillespie algorithm alone or in combination

FIG. 10. Number of (a) pentameric and (b) trimeric nuclei in lattice simulations without a partially formed capsid, for wild-type and Y132A mutant dimers. The template used for each analysis is shown on each graph. To be counted as a pentameric nucleus, three out of five dimers had to be classified in the cluster by the RMSD-fitting algorithm using the pentameric template shown. To be counted as a trimeric nucleus, all three dimers had to be classified in the cluster.

with fitting to experimental measurements.^{56,57} In addition, there are many possibilities for improvement of the model, as discussed below.

A majority of the simulations did not yield defect-free capsids, which may be a cause for concern or may be of scientific interest. As discussed below in detail, imperfect assembly could be an artifact of one or more aspects of our simulation protocol. However, it is also important to note that *in vitro* and *in vivo* capsid assemblies are stochastic processes in which defects may often occur: in fact, for other viruses, it has been noted experimentally that subpopulations of apparently intact capsids fail to mature fully^{75,76} which could result from the presence of defects. Further, cryo-electron microscopy capsid structure models are determined by averaging over the icosahedral symmetries⁷⁷ and hence are defect-free by construction. In improved future simulations, it will be interesting to observe the prevalence of defects to understand the degree to which they may be intrinsic to complex self-assembly processes.

Limitations

There are a number of limitations of this initial study. These include the energy function, the tabulation resolution and implementation, as well as the use of Monte Carlo simulation for dynamics.

Energy function

The Gō-like model used as the basis for the simulations presented here is still a relatively approximate, coarse-grained description of the interaction between subunits. The model's compromises in chemical accuracy can be expected to affect assembly pathways and capsid stability, perhaps significantly. In particular, the rapid formation of higher-order species mentioned above suggests that the interactions between capsid dimers may be too strong in our initial parametrization. The ultimate goal of providing mechanistic information consistent with experimental data likely will require a model incorporating the essentials of chemical specificity, such as local variation in hydrophobicity. Such a model, for example, would make it possible to assess the contributions to capsid formation from hydrophobic and electrostatic forces or to examine in more detail mutations such as the Y132A mutation.

It is possible to incorporate other more exact energy functions into the tables, such as models based on atomistic force fields. Once the tables have been constructed, in theory, it should take the same amount of time to look up an energy value regardless of the energy function used. (This was observed in our timing tests for the most part.) Consequently, it should be possible to use improved, perhaps atomistic energy functions without a decline in performance. The construction of such tables will be more time consuming than the construction of tables based on the Gō model, of course, because of the larger number of interatomic interactions that need to be calculated for each table entry. (Even with the relatively simple Gō model, the construction of the tables took a substantial amount of CPU time, about 3000 CPU-hours.) Nevertheless, we plan to pursue such tables as part of our future work.

Rigidity

It may be possible to overcome the constraint of dimer rigidity in future tabulation studies. In the study reported here, to permit tabulation of the interaction energy between dimers, it was assumed that each dimer was entirely rigid. Thus, internal conformational changes within each dimer were not possible in our implementation, although there is evidence that allosteric conformational changes within individual subunits play an important role in the assembly of the hepatitis B capsid.^{62,63} One way of including these conformational changes is to subdivide the subunits into smaller rigid fragments. This would reduce the computational speedup somewhat, since the rigid fragments would be smaller. Another possibility is to allow multiple reference geometries for each dimer and incorporate Monte Carlo moves which change the reference geometry associated with a given dimer. Either approach would require creating more tables and likely use more memory. However, the scaling behavior for memory use with fragment size is not completely trivial because each table for smaller fragments would require less memory at a fixed cutoff and resolution, as smaller fragments have a smaller six-dimensional relative configuration space.

Table resolution and memory

Figure 3 shows a clear dependence of structural fluctuations of the capsid on table resolution, with higher resolution tables resulting in simulations of the capsid that deviate less from the initial structure. However, the highest resolution tables, which yield the most stable capsid structures, require substantial memory (13.5 GB in this case), and since the tables are six-dimensional, increasing the resolution further will increase the size of the tables very rapidly. Also, the amount of memory needed will grow substantially with the complexity of the viral capsid or other multi-protein systems that will be simulated, since in general the number of tables needed will be proportional to the square of the number of distinct fragment types. There would be a similar increase in the number of tables if multiple reference geometries are used as described above. At the same time, memory capacity has increased rapidly over time and >100 GB is already common on inexpensive commodity clusters.

Nevertheless, memory will be a serious consideration for the tabulation strategy in the near future, and it will be useful to find ways to reduce the size of the tables. There are several possibilities. As suggested in our previous work with proteins,⁴⁵ the tables could be indexed according to $\cos \theta$ or $\cos \theta'$, which would reduce nonuniformity in the volume of table cells and thus give smaller tables for the same effective resolution. Variable resolution grids might also make it possible to represent the same function with less memory.

Fidelity of pathways and kinetics

The weighted ensemble method is notable for the ability to obtain accurate rates of conformational transitions^{78–81} and is statistically rigorous no matter what progress coordinate or binning strategy is used—even if (as here) the bin structure

changes over the course of the simulation.⁶⁷ Nevertheless, WE cannot overcome intrinsic limitations of the physics of trajectories: at short times *relative to the natural event durations of the system*, WE will yield trajectories that may not be representative of typical (slower) events; that is, the simulated events may be typical only of the very rare subset of events occurring on the time scale simulated.^{66,82,83} These time scales can be beyond the range accessible to simulation; for example, the natural transition event time t_b ranges from 2 to 200 μ s for the folding of small proteins.^{84,85} When the WE algorithm only samples events on a time scale faster than their natural event time, severely low weights may be obtained, leading to numerical difficulties. In future work, it will be important to probe the sensitivity of the observed pathways to the length of simulations to ensure that artifactual transients are not over-emphasized.

We also recognize that with finite sampling in WE simulations, the choice of progress coordinate and bin structure could affect the ability of the simulation to sample the relevant parts of configuration and path space. In the partial pathways obtained here, the central capsid is built up with one dimer at a time. This may be because the second progress coordinate is the distance between a single closest dimer and a position on the central capsid. Although WE simulation would eventually sample alternate pathways with the correct probabilities, sub-optimal progress coordinates could divert computing resources from those mechanisms. It follows that choices of progress coordinates more “agnostic” to the mechanism may be needed to obtain examples of other pathways, such as hierarchical pathways in which multiple partial capsids form and subsequently join together. We will explicitly explore this issue in future studies.

The choice of progress coordinate, analysis algorithm, and WE protocol may also be inhibiting our ability to simulate the complete process for other reasons. For example, the RMSD-based algorithm used for identifying partially formed capsids in the WE simulations does not consider which dimers in the simulation are physically in contact. As a result, when used with a complete capsid as a template and applied to simulations starting from a lattice without a preformed partial capsid, this algorithm can identify widely separated dimers as part of the same cluster, simply because they happen to have the same relative positions and orientations as a subset of the dimers in the capsid. Revising the algorithm to make better use of local structural information should make it easier to identify capsids in the early stages of their formation. Some of the final structures obtained, such as the ones in which the final dimer is identified as being part of the capsid without being fully attached or in which two layers of dimers partially “overlap” one another, suggest that the distance cutoff used with the RMSD algorithm is too large.

In addition, because the Monte Carlo method was used for the simulations presented here, it will not be possible to calculate physical rates for viral capsid assembly directly from them. In order to provide physical rates and time scales, the tabulation approach could be extended to work with molecular dynamics or Brownian dynamics. This would necessitate tabulating forces or interpolating between the values provided in the table to provide a continuously

differentiable potential. Since the tables are based on spherical coordinates, interpolation might be feasible using either spherical harmonics or spherical B-splines.⁸⁶ The resulting forces and torques would then be used with integration formulas suitable for rigid bodies.⁸⁷

Future plans

As described above, while a great deal of progress has been made in applying tabulation to construct a spatially realistic simulation of the late stages of hepatitis B viral capsid assembly, there are many improvements that will be made as part of our future work. These will include extending the simulations to try to provide a complete assembly pathway, constructing tables based on more realistic potentials such as atomistic force fields, devising new automated analysis algorithms selecting pathway-agnostic progress coordinates, identifying assembly mechanisms based on quantitative analysis of the simulations, and comparing these mechanisms with those obtained from other simulations and with available experimental data. If these goals can be accomplished, tabulation based simulations could provide spatially realistic simulations for the assembly of hepatitis B viral capsid and related systems, allowing insights into assembly not possible with previously reported simulations.

CONCLUSIONS

A previously developed alternative to traditional site-based coarse graining, namely, energy tabulation, has been applied for the first time to a multi-protein system in conjunction with the weighted ensemble (WE) multi-trajectory path sampling strategy. Encouragingly, the combination of approaches was able to yield trajectories describing the final stages of assembly of the hepatitis B viral capsid at significantly higher resolution than has previously been reported. Tabulation by itself reduced the cost of energy calculation by a factor exceeding 10^3 , with little further reduction in accuracy beyond the underlying energy function, and WE enabled the partial-assembly events which otherwise would have been inaccessible. Significant future work remains to be done, including especially the improvement of the underlying energy function and relaxation of built-in rigidity, more parsimonious use of memory via improved table structure, and exploration of WE coordinates that are demonstrably agnostic to assembly mechanism. Nevertheless, we believe the present work is a substantial step in validating the promise of an alternative, little-used strategy.

ACKNOWLEDGMENTS

We thank Dr. David Koes, Dr. Lillian Chong, Dr. Sundar Raman Subramanian, Dr. Ernesto Suarez, Dr. Ramu Anandakrishnan, Rory Donovan, Adam Pratt, Karl Debiec, and Josh Adelman for helpful discussions and technical assistance. We also thank the Department of Computational and Systems Biology and the Center for Simulation and Modeling at the University of Pittsburgh for computer time. This work was supported by NIH Grant Nos. P41-GM103712

and R01-GM115805, NSF Grant Nos. MCB-1119091 and CNS-1229064, and a Commonwealth Universal Research Enhancement Program grant from the Commonwealth of Pennsylvania Department of Health (No. SAP 4100062224).

APPENDIX: ALGORITHM FOR DETERMINATION OF PROGRESS COORDINATES FOR WE SIMULATIONS

1. The positions and orientations of the dimers in a fully formed capsid are obtained by RMSD-fitting the original reference geometry of each fragment to the corresponding chains in the crystal structure and transforming them through the symmetry transformations given in the PDB file. This produces a template corresponding to a fully formed capsid.
2. Given the positions and orientations of the fragments in a frame of the simulation, partially formed capsids are identified, and the net translation and rotation needed to bring the template into alignment with them are determined as follows:
 - (a) The identification of the first capsid is initialized by matching the first fragment in the frame to the first fragment of the same type in the template. The net translation and rotation for the first capsid are set so that this first fragment in the frame corresponds exactly to this matching fragment in the template.
 - (b) The identification of each capsid is updated via the following procedure:
 - (i) A “model” of the frame is created by rotating and translating the entire template through the current transformation for each currently identified partially formed capsid.
 - (ii) The intercenter distance between each fragment in the frame and each fragment of the same type in the model is determined and checked against a distance cutoff R . The angular distance between their orientations is checked against an angular cutoff α . If both distances are within their respective cutoffs, the pair is added to a list of correspondences between frame and template fragments. (The specific values of the cutoffs used in the present simulation are shown in Table I.)
 - (iii) The list of fragment pairs is sorted by intercenter distance. Working in order from least intercenter distance to greatest, each fragment from the frame is assigned to the corresponding partially formed capsid and fragment from the template, making sure not to assign the same fragment from the frame to two different fragments from the model or two fragments from the frame to the same fragment in the model.
 - (c) For each partially formed capsid, the current transformation is updated by RMSD fitting the fragment centers in the frame to the corresponding fragment centers in the template, using the updated assignment.
 - (d) The identification of capsids and the update of transformations are repeated until the assignments converge.

- (e) If any fragment from the frame remains unassigned, it is considered part of a new partially formed capsid and assigned to the first fragment of the same type in the template. The net translation and rotation for the first capsid are set so that this first fragment in the frame corresponds exactly to the matching fragment in the template. All of the above steps are then repeated to converge on an assignment including all fragments in the frame.
3. From the assignment, the largest partially formed capsid can be identified and the number of fragments in it is computed. This is the first progress coordinate. Any fragments in the template not mapped to this largest partial capsid are identified and the positions they would occupy are computed from the RMSD fit. The minimum distance between these “empty” positions and the positions of fragments in the frame that are not part of the largest capsid is also computed and becomes the second progress coordinate.

- ¹M. Levitt and A. Warshel, *Nature* **253**, 694 (1975).
- ²M. Levitt, *J. Mol. Biol.* **104**, 59 (1976).
- ³H. Taketomi, Y. Ueda, and N. Gö, *Int. J. Pept. Protein Res.* **7**, 445 (1975).
- ⁴Y. Ueda, H. Taketomi, and N. Gö, *Biopolymers* **17**, 1531 (1978).
- ⁵S. Tanaka and H. A. Scheraga, *Macromolecules* **9**, 945 (1976).
- ⁶S. Miyazawa and R. L. Jernigan, *Macromolecules* **18**, 534 (1985).
- ⁷A. Kolinski and J. Skolnick, *Proteins* **18**, 338 (1994).
- ⁸A. Kolinski and J. Skolnick, *Polymer* **45**, 511 (2004).
- ⁹M. M. Tirion, *Phys. Rev. Lett.* **77**, 1905 (1996).
- ¹⁰I. Bahar and A. J. Rader, *Curr. Opin. Struct. Biol.* **15**, 586 (2005).
- ¹¹P. A. Golubkhov and P. Ren, *J. Chem. Phys.* **125**, 064103 (2006).
- ¹²V. Tozzini, W. Rocchia, and J. A. McCammon, *J. Chem. Theory Comput.* **2**, 667 (2006).
- ¹³J.-W. Chu and G. A. Voth, *Biophys. J.* **93**, 3860 (2007).
- ¹⁴W. G. Noid *et al.*, *J. Phys. Chem. B* **111**, 4116 (2007).
- ¹⁵W. G. Noid *et al.*, *J. Chem. Phys.* **128**, 244115 (2008).
- ¹⁶W. G. Noid, *J. Chem. Phys.* **139**, 090901 (2013).
- ¹⁷A. Liwo, Y. He, and H. A. Scheraga, *Phys. Chem. Chem. Phys.* **13**, 16890 (2011).
- ¹⁸Y. Chebaro, S. Pasquali, and P. Derreumaux, *J. Phys. Chem. B* **116**, 8741 (2012).
- ¹⁹A. Davtyan *et al.*, *J. Phys. Chem. B* **116**, 8494 (2012).
- ²⁰S. Riniker, J. R. Allison, and W. F. van Gunsteren, *Phys. Chem. Chem. Phys.* **14**, 12423 (2012).
- ²¹M. G. Saunders and G. A. Voth, *Curr. Opin. Struct. Biol.* **22**, 144 (2012).
- ²²S. J. Marrink *et al.*, *J. Phys. Chem. B* **111**, 7812 (2007).
- ²³L. Monticelli *et al.*, *J. Chem. Theory Comput.* **4**, 819 (2008).
- ²⁴X. Periole *et al.*, *J. Chem. Theory Comput.* **5**, 2531 (2009).
- ²⁵M. Seo *et al.*, *J. Chem. Theory Comput.* **8**, 1774 (2012).
- ²⁶S. J. Marrink and D. P. Tieleman, *Chem. Soc. Rev.* **42**, 6801 (2013).
- ²⁷T. A. Wassenaar *et al.*, *J. Phys. Chem. B* **117**, 3516 (2013).
- ²⁸P. Kar *et al.*, *J. Chem. Theory Comput.* **9**, 3769 (2013).
- ²⁹J. F. Dama *et al.*, *J. Chem. Theory Comput.* **9**, 2466 (2013).
- ³⁰A. Davtyan *et al.*, *J. Chem. Theory Comput.* **10**, 5265 (2014).
- ³¹J. M. A. Grime and G. A. Voth, *J. Chem. Theory Comput.* **10**, 423 (2014).
- ³²Y. Levy *et al.*, *J. Mol. Biol.* **340**, 67 (2004).
- ³³Y. Levy *et al.*, *J. Mol. Biol.* **346**, 1121 (2005).
- ³⁴S. R. McGuffee and A. H. Elcock, *PLoS Comput. Biol.* **6**, e1000694 (2010).
- ³⁵C. T. Andrews and A. H. Elcock, *J. Chem. Theory Comput.* **10**, 5178 (2014).
- ³⁶T. Frembgen-Kesner *et al.*, *J. Chem. Theory Comput.* **11**, 2341 (2015).
- ³⁷N. Basdevant, D. Borgis, and T. Ha-Duong, *J. Phys. Chem. B* **111**, 9390 (2007).
- ³⁸Y. C. Kim and G. Hummer, *J. Mol. Biol.* **375**, 1416 (2008).
- ³⁹Y. C. Kim, R. B. Best, and J. Mittal, *J. Chem. Phys.* **133**, 205101 (2010).
- ⁴⁰M. Baaden and S. J. Marrink, *Curr. Opin. Struct. Biol.* **23**, 878 (2013).
- ⁴¹S. Takada, *Curr. Opin. Struct. Biol.* **22**, 130 (2012).
- ⁴²V. Tozzini, *Acc. Chem. Res.* **43**, 220 (2010).
- ⁴³M. G. Saunders and G. A. Voth, *Annu. Rev. Biophys.* **42**, 73 (2013).
- ⁴⁴S. Lettieri and D. M. Zuckerman, *J. Comput. Chem.* **33**, 268–275 (2012).

- ⁴⁵J. Spiriti and D. M. Zuckerman, *J. Chem. Theory Comput.* **10**, 5161–5177 (2014).
- ⁴⁶P. L. Freddolino *et al.*, *Structure* **14**, 437 (2006).
- ⁴⁷A. Arkhipov, P. L. Freddolino, and K. Schulten, *Structure* **14**, 1767 (2006).
- ⁴⁸G. Zhao *et al.*, *Nature* **497**, 643 (2013).
- ⁴⁹H. D. Nguyen, V. S. Reddy, and C. L. Brooks III, *Nano Lett.* **7**, 338 (2007).
- ⁵⁰H. D. Nguyen and C. L. Brooks III, *Nano Lett.* **8**, 4574 (2008).
- ⁵¹H. D. Nguyen, V. S. Reddy, and C. L. Brooks III, *J. Am. Chem. Soc.* **131**, 2606 (2009).
- ⁵²M. F. Hagan and D. Chandler, *Biophys. J.* **91**, 42 (2006).
- ⁵³M. R. Perkett and M. F. Hagan, *J. Chem. Phys.* **140**, 214101 (2014).
- ⁵⁴J. E. Baschek, H. C. R. Klein, and U. S. Schwarz, *BMC Biophys.* **5**, 22 (2012).
- ⁵⁵R. Schwartz *et al.*, *Biophys. J.* **75**, 2626 (1998).
- ⁵⁶L. Xie *et al.*, *Biophys. J.* **103**, 1545 (2012).
- ⁵⁷E. C. Dykeman, P. G. Stockley, and R. Twarock, *Proc. Natl. Acad. Sci. U. S. A.* **111**, 5361 (2014).
- ⁵⁸M. G. Mateu, *Arch. Biochem. Biophys.* **531**, 65 (2013).
- ⁵⁹C. Uetrecht *et al.*, *Nat. Chem.* **3**, 126 (2010).
- ⁶⁰A. Zlotnick *et al.*, *Biochemistry* **38**, 14644 (1999).
- ⁶¹P. Ceres and A. Zlotnick, *Biochemistry* **41**, 11525 (2002).
- ⁶²C. G. Alexander *et al.*, *Proc. Natl. Acad. Sci. U. S. A.* **110**, E2782 (2013).
- ⁶³C. Packianathan *et al.*, *J. Virol.* **84**, 1607 (2010).
- ⁶⁴S. A. Wynne, R. A. Crowther, and A. G. W. Leslie, *Mol. Cell* **3**, 771 (1999).
- ⁶⁵G. A. Huber and S. Kim, *Biophys. J.* **70**, 97 (1996).
- ⁶⁶B. W. Zhang, D. Jasnow, and D. M. Zuckerman, *Proc. Natl. Acad. Sci. U. S. A.* **104**, 18043 (2007).
- ⁶⁷B. W. Zhang, D. Jasnow, and D. M. Zuckerman, *J. Chem. Phys.* **132**, 054107 (2010).
- ⁶⁸M. Zwier *et al.*, *J. Chem. Theory Comput.* **11**, 800 (2015).
- ⁶⁹W. K. Kegel and P. v. d. Schoot, *Biophys. J.* **86**, 3905 (2004).
- ⁷⁰D. M. Zuckerman, *J. Phys. Chem. B* **108**, 5127 (2004).
- ⁷¹S. Matysiak and C. Clementi, in *Coarse-Graining of Condensed Phase and Biomolecular Systems*, edited by G. A. Voth (CRC Press, Boca Raton, 2009), p. 157.
- ⁷²P. Das, S. Matysiak, and C. Clementi, *Proc. Natl. Acad. Sci. U. S. A.* **102**, 10141 (2005).
- ⁷³A. Schug, C. Hyeon, and J. N. Onuchic, in *Coarse-Graining of Condensed Phase and Biomolecular Systems*, edited by G. A. Voth (CRC Press, Boca Raton, 2009).
- ⁷⁴E. A. Coutsiias, C. Seok, and K. A. Dill, *J. Comput. Chem.* **25**, 1849 (2004).
- ⁷⁵J. D. Perlmutter and M. F. Hagan, *Annu. Rev. Phys. Chem.* **66**, 217 (2015).
- ⁷⁶R. Tandon, E. S. Mocarski, and J. F. Conway, *Viruses* **7**, 899 (2015).
- ⁷⁷T. S. Baker, N. H. Olson, and S. D. Fuller, *Microbiol. Mol. Biol. Rev.* **63**, 862 (1999), available online at <http://mmbr.asm.org/content/63/4/862.short>.
- ⁷⁸R. M. Donovan *et al.*, *J. Chem. Phys.* **139**, 115105 (2013).
- ⁷⁹M. C. Zwier and L. T. Chong, *Curr. Opin. Pharmacol.* **10**, 745 (2010).
- ⁸⁰M. C. Zwier, J. W. Kaus, and L. T. Chong, *J. Chem. Theory Comput.* **7**, 1189 (2011).
- ⁸¹E. Suarez *et al.*, *J. Chem. Theory Comput.* **10**, 2658 (2014).
- ⁸²B. W. Zhang, D. Jasnow, and D. M. Zuckerman, *J. Chem. Phys.* **126**, 074504 (2007).
- ⁸³D. M. Zuckerman, *Statistical Physics of Biomolecules: An Introduction* (CRC Press, Boca Raton, FL, 2010).
- ⁸⁴H. S. Chung, J. M. Louis, and W. A. Eaton, *Proc. Natl. Acad. Sci. U. S. A.* **106**, 11837 (2009).
- ⁸⁵H. S. Chung and W. A. Eaton, *Nature* **502**, 685 (2013).
- ⁸⁶C. Jekeli, *Spline Representations of Functions on a Sphere for Geopotential Modeling* (The Ohio State University, Columbus, 2005).
- ⁸⁷E. Dickinson, S. A. Allison, and J. A. McCammon, *J. Chem. Soc., Faraday Trans. 2* **81**, 591 (1985).