HHS Public Access

Author manuscript

J Chem Theory Comput. Author manuscript; available in PMC 2018 November 14.

Published in final edited form as:

J Chem Theory Comput. 2017 November 14; 13(11): 5753–5765. doi:10.1021/acs.jctc.7b00840.

Hybrid All-Atom/Coarse-Grained Simulations of Proteins by Direct Coupling of CHARMM and PRIMO Force Fields

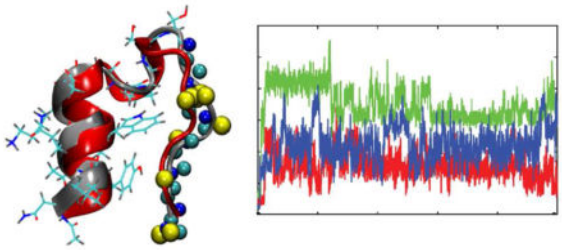
Parimal Kar and **Michael Feig***

Department of Biochemistry and Molecular Biology, Michigan State University East Lansing, Michigan 48824, United States

Abstract

Hybrid all-atom/coarse-grained (AA/CG) simulations of proteins offer a computationally efficient compromise where atomistic details are only applied to biologically relevant regions while benefitting from the speedup of treating the remaining parts of a given system at the CG level. The recently developed CG model, PRIMO, allows a direct coupling with an atomistic force field with no additional modifications or coupling terms and the ability to carry out dynamic simulations without any restraints on secondary or tertiary structures. A hybrid AA/CG scheme based on combining all-atom CHARMM and coarse-grained PRIMO representations was validated via molecular dynamics and replica exchange simulations of soluble and membrane proteins. The AA/CG scheme was also tested in the calculation of the free energy profile for the transition from the closed to the open state of adenylate kinase via umbrella sampling molecular dynamics method. The overall finding is that the AA/CG scheme generates dynamics and energetics that is qualitatively and quantitatively comparable to AA simulations while offering the computational advantages of coarse-graining. This model opens the door to challenging applications where high accuracy is required only in parts of large biomolecular complexes.

Graphical Abstract



*Corresponding author: feig@msu.edu, 603 Wilson Road, Room BCH 218, East Lansing, MI 48824, Phone: +1 (517) 432-7439, Fax: +1 (517) 353-9334.

SUPPORTING INFORMATION

Additional Figures S1 to S7 provide detailed RMSD and secondary structure and torsion angle time series. This information is available free of charge via the Internet at http://pubs.acs.org.

INTRODUCTION

Molecular dynamics (MD) simulations are frequently employed to investigate the structure and dynamics of biomolecular systems as a complement to experiments.[1, 2] There are numerous successful applications,[3–6] but conventional all-atom MD simulations of solvated bio-systems remain still very expensive, limiting the ability to fully match biological spatio-temporal scales.

One approach to addressing this limitation is to simplify the underlying system by means of coarse-graining (CG) approaches, in which several atoms are grouped into a single bead. A number of CG models for proteins with varying resolutions have been developed over the past years.[7, 8] However, the loss in chemical details often hinders quantitative studies of many complex biological processes, such as molecular recognition for which the explicit inclusion of atomistic details remains a must. Often such processes are localized to small portions of a given system so that multiscale all-atom/coarse-grained (AA/CG) approaches[9–15] could be used, where the biologically most relevant region of a system is treated at the atomic level of detail, while the rest of a given system is represented only at the CG level. Such a hybrid approach would preserve the advantages of both AA and CG approaches, i.e., providing accuracy where it matters while still reducing the computational cost.

AA/CG schemes have been proposed previously. One popular approach involves modeling an atomistic solute immersed in CG or dual-resolution AA/CG solvent[16–21] environments, mainly to improve the scalability of atomistic simulations. For instance, Shi and co-workers[22] have employed force-matching technique to simulate an atomistic ion channel in a CG membrane and in CG water. However, extensive parametrization was required to balance AA and CG interactions. Similarly, Orsi and co-workers[19, 23] have conducted dual-resolution MD simulations of atomistic proteins in CG water or CG membrane environments. In another similar effort, the PACE force field couples a united-atom protein model with MARTINI CG water and/or MARTINI CG lipids.[24–26] In PACE, cross-resolution parameters were optimized by targeting experimental thermodynamic quantities. On the other hand, Zacharias et al.[27] have developed a hybrid united atom/coarse-grained (UA/CG) model for proteins where the interactions of protein main chain sites are based on the GROMOS united atom force field. However, non-bonded interactions between side chains and between side- chain and main-chain sites are calculated at the level of a CG model using the knowledge-based ATTRACT[28] potential.

Other efforts have attempted to directly mix AA and CG models at the macromolecular level with combinations of AA models with the popular MARTINI[29] model being the most common approach. Rzepiela and co-workers[30] have described a scheme to combine MARTINI[29] and GROMOS 53a6[31] force fields by employing massless virtual interaction sites on relevant atomistic groups. This is conceptually similar to the center-of-mass particles employed in adaptive resolution hybrid schemes where there is no direct interaction between CG and AA atoms. Instead, the force acting on each virtual site is distributed to its constituent atoms weighted by their masses. In hybrid simulations with virtual sites, the unbalanced electrostatic screening from the CG water may induce abnormal behaviors. As a

remedy, Wassenaar and co-workers [32] attempted to mimic the electrostatic coupling using different relative dielectric constants with more advanced models. Following this scheme, Sokkar and co-workers[33] found that the conventional CG model (MARTINI) couples too strongly with the AA model and that it leads to complete unfolding of a test protein within a very short time. To circumvent this problem, they introduced position-restrained AA water layers in the hybrid simulation. However, free energy calculations suggest that the restrained solvent layer still results in an over-stabilization of the protein native structure, which would limit applications to the study of protein dynamics with this scheme. Recently, Zavadlav and co-workers[34] have reported an adaptive resolution scheme (AdResS)[14, 35–37] simulation of protein G in multiscale water, in which water molecules dynamically change resolution between an atomistic (SPC)[38] and coarse-grained (MARTINI) model. Zhang and Chen[39] reported a multiscale enhanced sampling (MSES) method where CG models are coupled with AA ones through restraint potentials in order to enhance the sampling of complex atomistic protein energy landscape and a similar approach was also followed by separate work by Kidera et al.[40, 41] (also called MSES) and Roux et al.[42] Finally, the Zuckerman group[43] has developed a mixed AA/CG scheme for proteins based on a library-based Monte Carlo (LBMC) approach[44], in which protein side chain configurations are pre-calculated and stored in libraries, while bonded interactions along the backbone are treated explicitly.

There are fewer attempts to apply intramolecular hybrid AA/CG models where different parts of a given protein are modeled at different resolutions. One such example is given by the Carloni group[15] where MD simulations of proteins were carried out using a mixed resolution model. In this approach, the AA region was modeled by the GROMOS96 force field with only polar hydrogens explicitly considered while a simplified Gō potential was used for the CG region. An interface region was introduced between the AA and CG regions in order to bridge the large discontinuity between fine-grained and CG descriptions. Solvent-protein interactions were only treated in terms of stochastic and frictional forces, proportional to the particle velocity and mass. The need for setting up the interface region and the use of a Gō potential to stabilize the CG part of the model limit practical applications of this approach.

A full integration of AA and CG models into a fully dynamic hybrid multi-scale scheme requires energetic and structural compatibility between the different representations for such a method to be successful. Ideally, the CG model should reflect a very similar energy landscape as the AA model but with reduced features to allow accelerated sampling at the CG level without deviating too far from conformations that are favorable at the AA level. We have recently developed the coarse-graining model PRIMO[45, 46] specifically with the goal of maximizing compatibility with the latest all-atom CHARMM[47] force field for proteins. PRIMO allows the reconstruction of all-atom models at negligible computational cost to accuracies of 0.1 Å[48] and its energies are highly correlated with those of CHARMM[45] despite a significant reduction in the degrees of freedom. These features are unique among the CG models proposed so far and make PRIMO an ideal candidate for a tight coupling with atomistic force fields in an AA/CG approach. Recently, Predeus and coworkers[49] have employed PRIMO and CHARMM[47] force fields simultaneously for the conformational sampling of Trp-cage and melittin peptides in the presence of crowder proteins. In that work, crowder proteins were described at the CG level while the proteins of interests were

modeled with the atomistic CHARMM force field. However, there were no direct bonded interactions between CG and AA parts.

The objective of the present work is to assess the suitability of the PRIMO force field for direct mixing with the atomistic CHARMM force field within the same molecule. We tested the resulting AA/CG model in dynamics simulations of globular and membrane proteins. Furthermore, we evaluated the application of the AA/CG method in quantitative estimates of potentials of mean force for a large-scale conformational change in *Escherichia coli* adenylate kinase (AdK) via umbrella sampling MD simulations.

The rest of the paper is organized as follows. First, we briefly review the PRIMO model and the proposed AA/CG scheme. Then the simulation protocols are described, followed by results and discussions.

METHODS

Multiscale Model

We have adopted a three-component multiscale modeling scheme where the protein is represented at the dual-resolution (AA/CG) level and the surrounding solvent environment is treated at the continuum level via the generalized Born with molecular volume (GBMV)[50] model and augmented with an atomic solvation term (ASP)[51]. The PRIMO model and the AA/CG scheme is described in more detail below.

PRIMO CG Model

The PRIMO model and force field are described in detail elsewhere[45]. Here, we briefly summarize the main features of PRIMO. The backbone in PRIMO is represented with N, C_α, and a combined carbonyl site (CO) placed at the geometric center of the carbonyl C and O atoms. This ensures preserving the backbone hydrogen bonding interactions. Non-glycine side chains are represented with one to five CG sites. The CG sites were chosen in such a way that an analytical reconstruction of AA model is possible with near-atomistic accuracies.[48] PRIMO uses virtual sites that are constructed on the fly from the CG interaction sites in order to maintain better bonding geometries. However, the virtual sites do not take part in non-bonded interactions. The PRIMO energy function (see Eq. 1) follows an all-atom-like physically motivated force field with additional terms for a combined generalized Born/atomic solvation parameter (GB/ASP) implicit solvent, an explicit angle- and distance-based hydrogen bonding interaction potential, and spline-based bonded potentials to maintain correct bond geometries at the CG level. The GB term is used with the GBMV model[50, 52] which also provides an approximation for atomic solvent-accessible surface areas (SASA) that are needed for the ASP term.

$$U^{\mathrm{PRIMO}}=\sum_{i=1}^{N_{\mathrm{bond}}} k_i^{\mathrm{bond}}\left(l_i-l_{i,0}\right)^2+\sum_{i=1}^{N_{\mathrm{angle}}} k_i^{\mathrm{angle}}\left(\theta_i-\theta_{i,0}\right)^2+\sum_{i=1}^{N_{\mathrm{bond\text{-}spline}}} s_i^{1D}\left(l_i^{1-2}\right)+\sum_{i=1}^{N_{\mathrm{angle\text{-}spline}}} s_i^{1D}\left(l_i^{1-3}\right)+$$
$$\sum_{i=1}^{N_{\mathrm{virtual\text{-}atom}}} k_i^{\mathrm{virtual\text{-}bond}}\left(l_i-l_{i,0}\right)^2+\sum_{i=1}^{N_{\mathrm{virtual\text{-}atom}}} k_i^{\mathrm{virtual\text{-}angle}}\left(\theta_i-\theta_{i,0}\right)^2+$$
$$\sum_{i=1}^{N_{\mathrm{torsion}}}\sum_{j=1}^{N_i^{\mathrm{mult}}} k_{i,\mathrm{j}}^{\mathrm{torsion}}\left(1+\cos\left(n_{i,j}\varphi_i-\varphi_{i,j,0}\right)\right)+\sum_{i=1}^{N_{\mathrm{torsion\text{-}spline}}} s_{1\mathrm{d}}^{\mathrm{i}}\left(l_{\mathrm{i}}^{1-4}\right)+$$
$$\sum_{i=1}^{N_{\mathrm{CMAP}}} s_{\mathrm{i}}^{2\mathrm{D}}\left(\varphi_{\mathrm{i}},\psi_{\mathrm{i}}\right)+\sum_{\mathrm{i}=1}^{N_{\mathrm{HBOND}}} s_{\mathrm{i}}^{2\mathrm{D}}\left(\cos\theta, l_{\mathrm{i}}^{\mathrm{N\text{-}CO}}\right)+$$
$$\sum_{i=1}^{N_{atom}-1}\sum_{\mathrm{j=i+1}}^{N_{\mathrm{atom}}} 4\varepsilon_{\mathrm{ij}}\left[\left(\frac{\sigma_{\mathrm{ij}}}{r_{\mathrm{ij}}}\right)^{12}-\left(\frac{\sigma_{\mathrm{ij}}}{r_{\mathrm{ij}}}\right)^{6}\right]+\sum_{\mathrm{i}=1}^{N_{\mathrm{atom}}-1}\sum_{\mathrm{j=i+1}}^{N_{atom}}\frac{q_{\mathrm{i}}q_{\mathrm{j}}}{4\pi\varepsilon_0 r_{\mathrm{ij}}}+\Delta G_{\mathrm{solv}}^{\mathrm{GB}}+\sum_{i=1}^{N_{\mathrm{atom}}}\gamma_{\mathrm{i}}\left(SASA\right)_{\mathrm{i}}$$

(1)

In the membrane version of PRIMO, PRIMO-M[53], the standard GB term is simply replaced by the HDGB model[54], and the SASA (solvent accessible surface area) term is scaled in a *z*-dependent fashion along the membrane normal.

Mixed AA/CG Interactions

In the AA/CG scheme, we set the representation level of each CG or AA particle at the beginning of the simulation and it remains unchanged throughout the course of the simulations. The interactions at the fine-grained or atomistic level were evaluated by the CHARMM36/CMAP[55–57] force field combined with the GBMV model[50] while CG interactions were calculated via PRIMO[45, 46]. The two force fields are combined directly, without any additional scaling factors, buffer regions, extra sites, or any other modifications of either the CHARMM or PRIMO models.[19] The transition between CHARMM and PRIMO was accomplished by switching resolution at the amino acid residue boundary. Since the all-atom backbone sites are mostly retained in PRIMO, this allows for a straightforward implementation of bonded terms across the carbonyl-amino backbone bond connecting two residues by mixing bonding terms from either the CHARMM36 or PRIMO parameter sets as appropriate (see Figure 1). More specifically, parameters corresponding to the bond (e.g., C-N1) and angle (C_α-C-N1, C-N1-CA1, O-C-N1, etc.) were taken from the CHARMM36 parameter set while the torsion parameters were taken either from PRIMO (e.g., C-N1-CA1-CO) or CHARMM36 (e.g., C_α-C-N1-CA1) force fields. Standard combination rules were applied for the mixed Lennard-Jones interactions. Electrostatic interactions between CHARMM and PRIMO regions were simply calculated according to the Coulomb law with the respective partial charges in the CHARMM and PRIMO force fields. MD simulations were performed with the combined Hamiltonian for the hybrid system with the potential energy: $U = U_{\mathrm{AA}} + U_{\mathrm{CG}} + U_{\mathrm{AA/CG}}$.

Both CHARMM and PRIMO have been validated with implicit solvent and GBMV was used without modification for both levels of resolution to obtain the electrostatic part of the solvation free energy. As described previously, PRIMO has an additional atom-type dependent ASP term based on atomic SASA values (see Eq. 1) that was only applied to PRIMO sites. For the atomistic sites we used a simple non-polar SASA-based term with a surface tension coefficient of 0.015 kcal/mol/Å^2.

Molecular Systems and Model Building

To validate the AA/CG mixed resolution method, we conducted unrestrained MD simulations of three proteins of different sizes and folding topologies (mainly α, mainly β, and mixed α/β), namely Trp-cage (PDB code: 1L2Y), SH3 domain (PDB code: 1SHG), and protein G (PDB code: 3GB1), respectively (see Table 1). In the case of the Trp-cage mini-protein, we represented the first 14 residues at the AA level, while the remaining six residues were modeled at the CG level. The spectrin SH3 domain is a compact β-barrel made of five antiparallel β-strands, with a long 19-residue interconnecting loop between the first two strands. We modeled the 19-residue loop region at the CG level while the remaining part was described at the AA level. Finally, protein G consists of four antiparallel β-strands and an α-helix. In this case, we choose to represent the helical region at the AA level while the β-strand regions were modeled at the CG level. The CG content for the three proteins varied between 30 to 63% (see Table 1). For illustration, the starting structure of the SH3 domain modeled at the AA/CG level is shown in Figure 2. While most of the attention here is on soluble proteins, we also included a limited test of the membrane protein bacteriorhodopsin (PDB code: 1QHJ) where we combined PRIMO-M[53] with the CHARMM36 force field.[47, 55, 56] For the membrane protein, we chose to represent residues 5 to 103 at the AA detail while the remaining residues were considered at the CG level. Although the experimental structure contains retinal linked to Lys216 via a Schiff base we did not include retinal here since parameters for ligands are not available at the PRIMO level. For the four test proteins, the CG regions were chosen without regard of possible biological questions since the main focus for studying these systems is a proof of concept of the AA/CG scheme. Instead, we only considered avoiding a change in resolution in the middle of secondary structure elements when partitioning the systems into AA and CG regions.

Finally, we applied the AA/CG model to study a large conformational change in the phosphotransferase enzyme adenylate kinase (AdK). AdK is composed of three main domains, the core (CORE, residues 1–29, 68–117, and 161–214), the ATP binding domain (LID, residues 118–160), and the NMP binding domain (NMP, residues 30−67).[58] It is well known that, upon ligand binding, AdK undergoes a transition from the inactive open conformation to the catalytically active closed conformation (see Figure 3)[59]. In the AA/CG scheme, the CORE domain was treated at the CG level while the LID and NMP domains were modeled at the AA level since the biological focus is on the dynamics of the LID and NMP domains.

Molecular Dynamics Simulations

All of the proteins were capped with acetyl (ACE) and *N*-methyl (NME) groups to avoid artifacts of interacting zwitterionic N- and C-termini sometimes seen in small proteins

simulated using implicit solvent. Although the PRIMO model allows longer time steps, the AA part requires an integration time step of 2.0 fs which was used for AA resolutions as well as the AA/CG model during production simulations. A multiple time-stepping algorithm could be implemented in the future to further increase the performance of the AA/CG method but was not considered here. The time step was set to 1.5 fs during the heating phases to maintain stable structures initially. The SHAKE[60] algorithm was used to constrain bonds involving hydrogen atoms. All the simulations were carried out at 300 K and the temperature was controlled using a Langevin thermostat with a friction coefficient of 10 ps^{-1} applied to all non-hydrogen atoms. In the AA/CG simulations, non-bonded interactions were cut off at 17 Å with a smooth switching to zero starting at 14 Å. However, in the case of bacteriorhodopsin, larger cutoffs of 16 and 18 Å at the beginning and end of the switching region, respectively, were used as suggested previously[53]. The non-bonded interaction list cutoff was set to 20 Å. Bacteriorhodopsin was initially placed in the membrane with its principal axis aligned with the bilayer normal (*z*-axis) and its center of mass at the origin of the membrane bilayer.

For comparison, we also carried out pure CHARMM (AA) and PRIMO (CG) simulations. In the reference AA simulations, we used the CHARMM36[47, 50, 55, 56] force field in combination with either explicit (TIP3P water) or implicit (GBMV) solvent, termed AA/TIP3P and AA/GBMV, respectively. In the explicit solvent simulations, long-range electrostatic interactions were treated by the Particle-Mesh Ewald (PME) summation method[61] and van der Waals interactions were truncated with a cutoff of 10 Å. The PRIMO force field was employed in the reference CG simulations. A similar protocol as for the AA/CG simulations was followed for both the pure AA and CG simulations. However, an integration time step of 4 fs was used for the pure PRIMO simulation. All the simulations at different resolutions were conducted without any restraints starting from the corresponding energy-minimized experimental crystal structures. In the case of dual-resolution simulations, for each test protein, five independent simulations of 200 ns were conducted while reference simulations with AA/GBMV and at the CG level were also performed for 200 ns. The membrane protein bacteriorhodopsin was simulated for 50 ns at AA/CG and CG resolutions. To obtain timing information, we also simulated bacteriorhodopsin in an explicit POPC lipid bilayer. This system was setup using CHARMM-GUI[62]. Trp-cage in explicit solvent was simulated via 25 x 200 ns simulations, each started from different initial velocities, to fully capture the highly dynamic native state ensemble. Other proteins were simulated over 200 ns in triplicate in explicit solvent.

Replica Exchange Simulations of Trp-cage

In order to characterize and compare the conformational sampling of Trp-cage with the CG and AA/GBMV, and AA/CG models, temperature replica exchange molecular dynamics (REMD) simulations were carried out. All of the REMD simulations were started from the experimental structure. Eight to sixteen replicas were employed to span a temperature range of 300 – 500 K. Exchange moves between two consecutive replicas were attempted every 5 ps with an acceptance ratio of ~30%. In all three REMD simulations, each replica was simulated for at least 10 ns and two sets of REMD simulations were carried out for the AA/CG model.

Umbrella Sampling of Adenylate Kinase Loop Dynamics

Previous studies have suggested that the center of mass distance between the LID and CORE domains varies between 20.5 for the closed conformation to 29.5 Å for the open conformation.[63] A free energy profile as a function of the LID-CORE distance was generated here via umbrella sampling by varying the distance between 19 to 32 Å at steps of 0.5 Å. A force constant of 10 kcal mol^{-1} $Å^{-2}$ was employed to harmonically restrain the domains in each umbrella window at the respective center of mass distances. Production simulations in each window were carried out for 10 ns, much longer than previous umbrella sampling of 500 ps to 2 ns per window for this system.[58, 63, 64] The last 8 ns in each window were analyzed, and the weighted histogram analysis method (WHAM)[65] was used to generate a combined potential of mean force (PMF) along the one-dimensional reaction coordinate. The same protocol was followed for the AA and CG umbrella sampling MD simulations.

Simulation and Analysis

All of the AA, CG, and mixed AA/CG simulations were carried out using version c38a2 of the CHARMM macromolecular modeling package[47] where the PRIMO model is implemented. The MMTSB (Multiscale Modeling Tools for Structural Biology) Tool Set[66] in combination with CHARMM was employed for analysis. MSMBuilder[67], version 3.8, via the python API and custom-written scripts was used for principal component analysis. For the analysis of the CG and AA/CG simulations, we reconstructed AA models for the CG parts using a previously described trajectory reconstruction procedure that achieves better than 0.1 Å accuracy for heavy atoms.[45, 46, 48] All of the analysis presented here was done based on the all-atom reconstructed trajectories.

Availability

In order to run the AA/CG model described here, a recent version of the CHARMM macromolecular modeling package[47] is needed along with standard all-atom CHARMM force field files and PRIMO force field files. The PRIMO force field is available from the authors upon request in a package that also includes utility scripts to facilitate setup and analysis of PRIMO CG models as well as AA/CG hybrid models.

RESULTS

Stable AA/CG MD Simulation of Globular Soluble Proteins

As a first test, we investigated whether the AA/CG scheme can be used to run stable MD simulations of different soluble proteins. A set of three soluble proteins with 20 to 62 amino acids and different topologies was examined. We analyzed the performance of AA/CG simulations with respect to the thermodynamic stability and dynamic properties of the proteins in comparison with results from pure AA or CG simulations as well as experiments.

In the simulations, the proteins generally reach a stable conformation within the first 50 ns, and C_α root mean square deviation (RMSD) values are generally kept below 4 Å for the rest of the simulations when the AA/CG scheme is used (see time series for proteins G and the SH3 domain in Fig. S1). However, one copy of SH3 dynamically unfolded and refolded

towards larger RMSD values (see Fig. S1). Trp-cage also explored partially unfolded conformations during the simulations (see below). Average RMSD values varied between 2.42 and 3.71 Å and were lower, 1.70 to 2.75 Å when ensemble-averaged structures were compared with the experimental structures. Such deviations are somewhat larger than what is obtained in the AA/TIP3P (0.55–1.19 Å) and AA/GBMV (0.79–1.49 Å) simulations but they are lower than what the CG model alone provides for this set of proteins (2.94–4.35 Å). Conformation-averaged structures with the AA/CG model are shown in Figure 4 in comparison with the experimental structures. It can be seen that the native structures are overall well maintained with only minor deviations that appear to be more prominent in the regions modeled at the CG level (shown in blue in Fig. 4).

Average radii of gyration (R_g) were similar (see Table 2), but consistently larger than the experimental values and the averages obtained from the AA/TIP3P simulations indicating a slight overall expansion of the structures with the AA/CG model. Interestingly, using implicit solvent vs. explicit solvent with just the AA model appears to have the opposite effect (see Table 2). This suggests that replacing explicit solvent with implicit solvent may restrict conformational sampling to more compact states while the use of the hybrid model leads to broader conformational sampling where the proteins are at least as flexible in the AA/CG simulations as in explicit solvent. A similar trend was observed as well in previous hybrid simulations[68]. The tendency towards more expanded states with the AA/CG model is also reflected in increased solvent accessible surface areas (SASA) as shown in Table 3 although molecular volumes are more similar. The increase in SASA is seen for both the AA and CG regions and is larger than using either the AA or CG models alone (see Table 3). This suggests that there may be weakened hydrophobic packing interactions in the hybrid model between the AA and CG regions that lead to a limited expansion of the structures.

We further characterized the conformational sampling of the three test systems via two-dimensional potentials of mean force (PMF) as a function of RMSD and radius of gyration (Figure 5) as well as the first principal components (Figure 6). The principal components were obtained from a combined ensemble that included simulations at all model resolutions. It is apparent that that the sampling with the AA/CG model is broader, involving more extended states compared to the AA/TIP3P simulation. In all cases the, native state is within the low-lying areas of the energy landscapes with the AA/CG model and the overall free energy minima are close to the native state (see Fig. 5). The agreement with the AA/TIP3P sampling is especially good for the Trp-cage system where the overall shape of the energy landscape matches well with between AA/CG and AA/TIP3P with the main minimum located between 1 and 2 Å RMSD as in previous simulations of Trp-cage in both explicit and implicit solvents[45]. While the sampling with the AA/CG is also similar to AA/TIP3P for protein G and the SH3 domain when projected onto RMSD and radius of gyration, there are differences in the projections on the principal components, especially for SH3, that suggest that different unfolded states are visited (see Fig. 6). Again, the AA/GBMV ensembles show more restricted sampling whereas the ensembles sing PRIMO alone lead to ensembles that deviate more significantly from the native structure and the sampling seen with AA/TIP3P (see Figs. 5 and 6). Overall, the finding is that the AA/CG model generates conformational ensembles that overlap significantly with the sampling generated with explicit solvent but extend more broadly to include partially unfolded states. Broader conformational sampling

while still maintaining similar relative thermodynamics of native states is one goal of successful CG models. However, the resulting increased conformational flexibility seen here is in contrast to the model by Orsi and co-workers[19] where the flexibility decreased for the hybrid systems compared to AA simulations.

We further analyzed how the structural stability varies between the regions represented at AA and CG levels. The per-residue RMSD values for both C_α and side chains are shown in Figure 7. With the AA/CG model, structural deviations are somewhat larger for the CG regions than for the AA regions in protein G and SH3 and there is also a moderate increase in RMSD for the AA regions with values that are generally intermediate between pure AA and CG models. However, for Trp-cage, the deviations are more similar to the AA models (see Fig. 7). This indicates that the AA/CG model may affect not just the sampling of the CG region but also to some extent of the AA region. The trends observed based on the C_α atoms are generally reflected in the side chains but with slightly larger RMSD values. Figure 7 also shows the root-mean-square-fluctuations (RMSF) of C_α atoms. Again, the structural fluctuations with the AA/CG model are larger than with the AA models and less than with the pure CG model, except for the loop region between residues 12 and 29 of the SH3 domain, where there is significantly more dynamics with the AA/CG model due to the partial unfolding in one of the simulations. The structural fluctuations are generally larger with the AA/CG model, not just in the regions represented at the CG models, again indicating that the partial CG representation affects the sampling of the entire structure.

We further analyzed the preservation of secondary structures and the sampling of backbone φ/ψ and χ_1 side chain torsion angles. Overall secondary structures such as helical and β-strand regions were generally well preserved between AA/CG and AA/TIP3P simulations based on hydrogen-bonding analysis via DSSP[69] (see Figs S2 and S3). There were only minor variations at the edges of secondary structure elements that may be expected in dynamic simulations of flexible systems. The sampling of backbone φ/ψ torsions was also overall very similar between AA/CG and AA/TIP3P simulations (see Figs S4 and S5 and Fig. 8). Only some residues showed significant deviations in average backbone torsion angles between AA/CG and AA/TIP3P simulations and the torsion angles found in the experimental structures (see Fig. 8). The few significant deviations were seen mostly in the CG regions, including some residues at the interface between the AA and CG regions (for example in protein G and Trp-cage) but this analysis does not reveal significant problems with maintaining torsional sampling across the AA/CG boundaries. The sampling of side chains, characterized here via the χ_1 torsions showed variations between the simulations and the experimental side chain values (see Fig. 8) and also between AA/CG and AA/TIP3P simulations (see Fig. S6 and S7). However, the general features were mostly similar and strong boundary effects were not apparent suggesting that the overall sampling of side chain conformations is also comparable between the AA/CG and AA/TIP3P models.

AA/CG MD Simulation of Bacteriorhodopsin in an Implicit Membrane

We also carried out limited testing of whether PRIMO-M, the membrane version of PRIMO is suitable for AA/CG simulations. We conducted AA/CG MD simulation of bacteriorhodopsin (PDB code: 1QHJ) by combining PRIMO-M and CHARMM36 with the

HDGB implicit membrane model. The RMSD plot shown in Fig. 9 suggests that the AA/CG simulation is stable and comparable to previously published results with PRIMO[53]. The average C_α RMSD obtained from CHARMM/PRIMO simulation was 3.2 Å while the CG simulation (PRIMO-M) yielded an average C_α RMSD of 4.0 Å. C_α RMSDs of the average structure over the entire trajectory were closer to the native with 2.8 and 3.6 Å for CHARMM/PRIMO-M and PRIMO-M simulations, respectively. The same protein was simulated earlier by Tanizaki and Feig[54] using the CHARMM22/CMAP force field with HDGB [47, 70] where an average RMSD of 2.0 Å was reported from a 10 ns simulation. These results suggest that it may be possible to transfer the hybrid scheme to the membrane environment and thus can be employed to investigate the dynamic behaviors of membrane-interacting peptides and proteins. However, this will need to be confirmed in a more thorough later investigation of how the AA/CG hybrid scheme performs for membrane environments.

Free Energy of Large Conformational Transition in Adenylate Kinase

As a final test we explored the ability of the AA/CG approach to reproduce free energy profiles for large conformational transitions. We chose the well-studied open-to-close transition in adenylate kinase (AdK) as the test case. Following previous studies, we applied umbrella sampling combined with the WHAM algorithm to generate PMFs along the distance between the CORE and LID domains as the reaction coordinate. Figure 10 displays the resulting free energy profile of the LID motion computed with the AA/GBMV, AA/CG, and CG models. In all cases, the free energy profile of the LID motion is characterized by a broad single-well as observed by Arora and Brooks[58] and Song and Zhu[71] consistent with broad conformational space exploration of the LID domain due to its flexible nature. It should be noted here that Arora and Brooks[58] used the CHARMM22/CMAP[47, 72] force field with GBMV[50] while Song and Zhu[71] utilized the CHARMM36[57] force field with TIP3P water molecules. The free energy profile also indicates that the LID domain explores conformations ranging from the crystallographic open to the closed structure, which is in agreement with experiments.[73–75] We note that our simulations suggest that the *apo*-closed state is not a metastable stable, and the *apo*-open state is energetically more favorable than the closed state for unligated AdK. The free energy difference between the open and closed states estimated from CHARMM/PRIMO was found to be ~8 kcal/mol, closer to the result with CHARMM (~9.5 kcal/mol) than with PRIMO (~6.5 kcal/mol). Our AA and AA/CG estimates of the relative energy between the open and closed state are comparable to other calculations that found 6–13 kcal/mol with different methods and force fields.[7671775878] However, we note that our findings are in contrast to the results by Lou and Cukier[79] that found the open state to be energetically unfavorable using an Amber force field.

Most relevant for this study is the excellent qualitative and quantitative agreement between the PMFs generated with the AA/CG and AA models while the CG results only approximately reproduce the AA results. Hence, in this example, PRIMO alone can qualitatively capture the dynamic behavior of AdK while the AA/CG model also quantitatively reproduces the reference AA simulations.

Computational Efficiency of the AA/CG Model

Hybrid models are attractive because of their computational efficiency and accuracy. The speedup of a hybrid AA/CG model intrinsically depends on the relative sizes of AA and CG components present in the system. The single-core computational performance with the AA/CG, AA/GBMV, and AA/TIP3P models (POPC lipids were included for the membrane system) was compared for all five proteins (see Table 4). The AA/CG scheme achieves a speedup in the range of 1.4 to 2.3 compared to AA/GBMV while there is a two- to five-fold increase in speed compared to explicit water simulations. For comparison, using the PRIMO CG model alone is about ten times faster than AA/GBMV while speed-up by factors of 10 to 20 could be achieved with PRIMO compared to explicit solvent[45]. We are focusing on comparing single-core performances here because the parallel efficiency of explicit and implicit solvent and AA and CG simulations varies.

The much larger speedup of PRIMO alone vs. the AA/CG model is in part due to a longer time step that was used with the CG model. A standard AA time step of 2 fs was used here for the AA/CG simulations, but it is in principle possible to employ a multiple time-step approach, where the CG part would be propagated with a longer time step[13, 80] to achieve a greater speedup. Furthermore, applications that involve larger molecules may consist of only a small region simulated at AA resolution while the majority of the system would be modeled at the CG level. In such a case, and assuming the use of a multiple time step algorithm, the computational cost of the AA/CG model could become similar to using just a CG model.

As the timing results also show, the AA/GBMV simulations become almost as expensive as the explicit solvent models for the larger systems (adenylate kinase and bacteriorhodopsin). As noted before, this is a consequence of the relatively high cost of the GBMV implicit solvent model that is used here.[81] Explicit solvent simulations further benefit from better parallelization schemes[82, 83] and GPU acceleration[84–86]. However, GPU acceleration is also possible for GB models[87] and the replacement of the very expensive GBMV model by better-performing alternative GB implementations and other implicit solvent models, especially when combined with PRIMO or the PRIMO part of an AA/CG model remains a promising avenue for increased computational efficiency.

DISCUSSION AND CONCLUSION

The results presented here demonstrate that a direct-coupling AA/CG hybrid scheme based on the CHARMM and PRIMO models is feasible and can produce conformational sampling that is overall comparable to fully atomistic simulations. The AA/CG scheme was also found to not only provide qualitatively similar results to AA models but also quantitatively match the energetic profile for the open/close transition in adenylate kinase. We expect that AA/CG schemes would be especially useful for detailed mechanistic studies of dynamic elements in the context of very large complexes, such as the trigger loop in RNA polymerase II that required substantial computational resources in previous work from our lab using fully atomistic representations.[6]

A key factor in the success of the introduced AA/CG scheme is the use of the PRIMO CG model that by itself is already able to describe the dynamics of peptides and proteins in an approximate fashion without requiring any constraints and that was parameterized based on the same AA force field (CHARMM36) that is used here in combination with PRIMO. Similar functional forms of PRIMO and CHARMM36 make a direct combination possible without the need of a buffer or coupling region, and in fact the AA/CG model that was tested here did not involve any further parameterization whatsoever. The results presented here indicate, though, that some tuning of non-bonded interactions could improve the model, in particular to address what appear to be reduced hydrophobic packing interactions that lead to slightly more expanded structures. The possibility of reconstructing structures at atomistic detail from the CG model using analytical functions at negligible cost would also allow a more elaborate AA/CG scheme where instead of direct mixing as done here, the CG model would see a reduced CG counterpart for residues in the AA region immediately adjacent to the CG region and the AA model would see a reconstructed atomistic counterpart for residues in the CG region next to the AA region. Such a scheme where resolutions are extended across the interface region to minimize boundary effects may be more similar in spirit to many QM/MM implementations. Tuning the direct-mixing protocol as well as a more integrated double-resolution interface model will be explored in future work.

An interesting feature of our AA/CG scheme is that the reliance on strongly physically motivated energy terms allows broader transferability than with most CG models. As an example, we are demonstrating that simply by switching the implicit solvent model to an implicit membrane version, it is possible to simulate membrane-bound systems. While we only show preliminary results here, we will also test further in future work how AA/CG hybrid models may be applicable for membrane environments, including coupling to DHDGB model developed by us[88] to include membrane flexibility within the AA/CG scheme.

Ultimately, our goal is to develop a comprehensive multi-scale approach for modeling biomolecules in their native environments that would involve atomistic and coarse-grained levels as well as implicit solvent and elastic continuum models to cover different parts at different resolutions. We believe that such complex multi-tiered models are ultimately needed to simulate biological systems in highly complex cellular environments over biologically relevant time scales. We believe that the AA/CG scheme presented here is a first step in that direction.

The relatively high resolution of the PRIMO CG model and the use of the expensive GBMV implicit solvent model so far result in only modest computational gains. However, after establishing the proof of principle that fully dynamic AA/CG schemes can provide a reasonable substitute for full AA models, we will now also focus our attention on improving performance aspects by exploring alternative GB implementations, implementing multiple time step algorithms that would propagate the CG part with longer time steps, and optimizing parallel and GPU efficiencies.

In conclusion, we have presented the CHARMM/PRIMO AA/CG scheme that compares favorably with atomistic simulations both qualitatively and quantitatively while offering the

computational advantages of coarse-graining, both in terms of reduced computational costs and broader sampling. The model appears to be transferable to membrane environments, and it is in principle applicable to any protein system without requiring any structural restraints that are commonly needed in other CG models. In principle, this opens up a wide variety of mechanistic questions that could be studied with the AA/CG scheme introduced here but further assessment is needed to establish how this method performs for a larger set of systems and with different partition schemes between the AA and CG regions in terms of accuracy and computational efficiency.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

PK thanks Dr. Vahid Mirjalili for useful discussions. MF thanks Dr. Lim Heo for providing principal component analysis scripts. Funding from the National Institutes of Health (GM084953 and GM092949) and computer time at NSF XSEDE resources (TG-MCB090003) are acknowledged.

References

1. Karplus M, McCammon JA. Molecular Dynamics Simulations of Biomolecules. Nat Struct Biol. 2002; 9:646–652. [PubMed: 12198485]
2. van Gunsteren WF, Bakowies D, Baron R, Chandrasekhar I, Christen M, Daura X, Gee P, Geerke DP, Glattli A, Hunenberger PH, Kastenholz MA, Ostenbrink C, Schenk M, Trzesniak D, van der Vegt NFA, Yu HB. Biomolecular Modeling: Goals, Problems, Perspectives. Angew Chemie Int Ed. 2006; 45:4064–4092.
3. Harada R, Sugita Y, Feig M. Protein Crowding Affects Hydration Structure and Dynamics. J Am Chem Soc. 2012; 134:4842–4849. [PubMed: 22352398]
4. Shaw DE, Maragakis P, Lindorff-Larsen K, Piana S, Dror RO, Eastwood MP, Bank JA, Jumper JM, Salmon JK, Shan YB, Wriggers W. Atomic-Level Characterization of the Structural Dynamics of Proteins. Science. 2010; 330:341–346. [PubMed: 20947758]
5. Sharma M, Predeus AV, Kovacs N, Feig M. Differential Mismatch Recognition Specificities of Eukaryotic MutS Homologs, MutSα and MutSβ. Biophys J. 2014; 106:2483–2492. [PubMed: 24896128]
6. Wang BB, Predeus AV, Burton ZF, Feig M. Energetic and Structural Details of the Trigger-Loop Closing Transition in RNA Polymerase II. Biophys J. 2013; 105:767–775. [PubMed: 23931324]
7. Kar P, Feig M. Recent Advances in Transferable Coarse-Grained Modeling of Proteins. Adv Prot Chem Struct Biol. 2014; 96:143–180.
8. Ingolfsson HI, Lopez CA, Uusitalo JJ, de Jong DH, Gopal SM, Periole X, Marrink SJ. The Power of Coarse Graining in Biomolecular Simulations. Wiley Interdiscp Rev: Comput Mol Sci. 2014; 4:225–248.
9. Tozzini V. Multiscale Modeling of Proteins. Acc Chem Res. 2010; 43:220–230. [PubMed: 19785400]
10. Ayton GS, Noid WG, Voth GA. Multiscale Modeling of Biomolecular Systems: in Serial and in Parallel. Curr Opin Struct Biol. 2007; 17:192–198. [PubMed: 17383173]
11. Kamerlin SCL, Warshel A. Multiscale Modeling of Biological Functions. Phys Chem Chem Phys. 2011; 13:10401–10411. [PubMed: 21526232]
12. Peter C, Kremer K. Multiscale Simulation of Soft Matter Systems - From the Atomistic to the Coarse-Grained Level and Back. Soft Matter. 2009; 5:4357–4366.
13. Di Pasquale N, Gowers RJ, Carbone P. A Multiple Time Step Scheme for Multiresolved Models of Macromolecules. J Comput Chem. 2014; 35:1199–1207. [PubMed: 24676734]

14. Nielsen SO, Bulo RE, Moore PB, Ensing B. Recent Progress in Adaptive Multiscale Molecular Dynamics Simulations of Soft Matter. Phys Chem Chem Phys. 2010; 12:12401–12414. [PubMed: 20734007]
15. Neri M, Anselmi C, Cascella M, Maritan A, Carloni P. Coarse-Grained Model of Proteins Incorporating Atomistic Detail of the Active Site. Phys Rev Lett. 2005; 95:218102. [PubMed: 16384187]
16. Zavadlav J, Melo MN, Cunha AV, de Vries AH, Marrink SJ, Praprotnik M. Adaptive Resolution Simulation of MARTINI Solvents. J Chem Theory Comput. 2014; 10:2591–2598. [PubMed: 26580779]
17. Fuhrmans M, Sanders BP, Marrink SJ, de Vries AH. Effects of Bundling on the Properties of the SPC Water Model. Theor Chem Acc. 2010; 125:335–344.
18. Darre L, Tek A, Baaden M, Pantano S. Mixing Atomistic and Coarse Grain Solvation Models for MD Simulations: Let WT4 Handle the Bulk. J Chem Theory Comput. 2012; 8:3880–3894. [PubMed: 26593029]
19. Orsi M, Ding W, Palaiokostas M. Direct Mixing of Atomistic Solutes and Coarse-Grained Water. J Chem Theory Comput. 2014; 10:4684–4693. [PubMed: 26588159]
20. Riniker S, van Gunsteren WF. Mixing Coarse-Grained and Fine-Grained Water in Molecular Dynamics Simulations of a Single System. J Chem Phys. 2012; 137:044120. [PubMed: 22852610]
21. Renevey A, Riniker S. Improved Accuracy of Hybrid Atomistic/Coarse-Grained Simulations Using Reparametrised Interactions. J Chem Phys. 2017; 146:124131. [PubMed: 28388132]
22. Shi Q, Izvekov S, Voth GA. Mixed Atomistic and Coarse-Grained Molecular Dynamics: Simulation of a Membrane-Bound Ion Channel. J Phys Chem B. 2006; 110:15045–15048. [PubMed: 16884212]
23. Orsi M, Noro MG, Essex JW. Dual-Resolution Molecular Dynamics Simulation of Antimicrobials in Biomembranes. J R Soc, Interface. 2011; 8:826–841. [PubMed: 21131331]
24. Han W, Schulten K. Further Optimization of a Hybrid United-Atom and Coarse-Grained Force Field for Folding Simulations: Improved Backbone Hydration and Interactions between Charged Side Chains. J Chem Theory Comput. 2012; 8:4413–4424. [PubMed: 23204949]
25. Wan CK, Han W, Wu YD. Parameterization of PACE Force Field for Membrane Environment and Simulation of Helical Peptides and Helix-Helix Association. J Chem Theory Comput. 2012; 8:300–313. [PubMed: 26592891]
26. Ward MD, Nangia S, May ER. Evaluation of the Hybrid Resolution PACE Model for the Study of Folding, Insertion, and Pore Formation of Membrane Associated Peptides. J Comput Chem. 2017; 38:1462–1471. [PubMed: 28102001]
27. Zacharias M. Combining Coarse-Grained Nonbonded and Atomistic Bonded Interactions for Protein Modeling. Proteins. 2013; 81:81–92. [PubMed: 22911567]
28. Zacharias M. Protein-Protein Docking with a Reduced Protein Model Accounting for Side-Chain Flexibility. Protein Sci. 2003; 12:1271–1282. [PubMed: 12761398]
29. Monticelli L, Kandasamy SK, Periole X, Larson RG, Tieleman DP, Marrink SJ. The MARTINI Coarse-Grained Force Field: Extension to Proteins. J Chem Theory Comput. 2008; 4:819–834. [PubMed: 26621095]
30. Rzepiela AJ, Louhivuori M, Peter C, Marrink SJ. Hybrid Simulations: Combining Atomistic and Coarse-Grained Force Fields Using Virtual Sites. Phys Chem Chem Phys. 2011; 13:10437–10448. [PubMed: 21494747]
31. Oostenbrink C, Villa A, Mark AE, van Gunsteren WF. A Biomolecular Force Field Based on the Free Enthalpy of Hydration and Solvation: The GROMOS Force-Field Parameter Sets 53A5 and 53A6. J Comput Chem. 2004; 25:1656–1676. [PubMed: 15264259]
32. Wassenaar TA, Ingolfsson HI, Priess M, Marrink SJ. Mixing MARTINI: Electrostatic Coupling in Hybrid Atomistic-Coarse-Grained Biomolecular Simulations. J Phys Chem B. 2011; 117:3516–3530.
33. Sokkar P, Choi SM, Rhee YM. Simple Method for Simulating the Mixture of Atomistic and Coarse-Grained Molecular Systems. J Chem Theory Comput. 2013; 9:3728–3739. [PubMed: 26584124]

34. Zavadlav J, Melo MN, Marrink SJ, Praprotnik M. Adaptive Resolution Simulation of an Atomistic Protein in MARTINI Water. J Chem Phys. 2014; 140:054114. [PubMed: 24511929]
35. Lyman E, Ytreberg FM, Zuckerman DM. Resolution Exchange Simulation. Phys Rev Lett. 2006; 96:028105. [PubMed: 16486650]
36. Praprotnik M, Delle Site L, Kremer K. Multiscale Simulation of Soft Matter: From Scale Bridging to Adaptive Resolution. Annu Rev Phys Chem. 2008; 59:545–571. [PubMed: 18062769]
37. Fritsch S, Poblete S, Junghans C, Ciccotti G, Delle Site L, Kremer K. Adaptive Resolution Molecular Dynamics Simulation Through Coupling to an Internal Particle Reservoir. Phys Rev Lett. 2012; 108:170602. [PubMed: 22680848]
38. Berweger CD, van Gunsteren WF, Müller-Plathe F. Force-Field Parametrization by Weak-Coupling - Reengineering SPC Water. Chem Phys Lett. 1995; 232:429–436.
39. Zhang WH, Chen JH. Accelerate Sampling in Atomistic Energy Landscapes Using Topology-Based Coarse-Grained Models. J Chem Theory Comput. 2014; 10:918–923. [PubMed: 26580171]
40. Moritsugu K, Terada T, Kidera A. Disorder-to-Order Transition of an Intrinsically Disordered Region of Sortase Revealed by Multiscale Enhanced Sampling. J Am Chem Soc. 2012; 134:7094–7101. [PubMed: 22468560]
41. Moritsugu K, Terada T, Kidera A. Scalable Free Energy Calculation of Proteins via Multiscale Essential Sampling. J Chem Phys. 2010; 133:224105. [PubMed: 21171681]
42. Tempkin JOB, Qi B, Saunders MG, Roux B, Dinner AR, Weare J. Using Multiscale Preconditioning to Accelerate the Convergence of Iterative Molecular Calculations. J Chem Phys. 2014; 140:184114. [PubMed: 24832260]
43. Mamonov AB, Lettieri S, Ding Y, Sarver JL, Palli R, Cunningham TF, Saxena S, Zuckerman DM. Tunable, Mixed-Resolution Modeling Using Library-Based Monte Carlo and Graphics Processing Units. J Chem Theory Comput. 2012; 8:2921–2929. [PubMed: 23162384]
44. Mamonov AB, Bhatt D, Cashman DJ, Ding Y, Zuckerman DM. General Library-Based Monte Carlo Technique Enables Equilibrium Sampling of Semi-atomistic Protein Models. J Phys Chem B. 2009; 113:10891–10904. [PubMed: 19594147]
45. Kar P, Gopal SM, Cheng YM, Predeus AV, Feig M. PRIMO: A Transferable Coarse-Grained Force Field for Proteins. J Chem Theory Comput. 2013; 9:3769–3788. [PubMed: 23997693]
46. Gopal SM, Mukherjee S, Cheng YM, Feig M. PRIMO/PRIMONA: A Coarse-Grained Model for Proteins and Nucleic Acids that Preserves Near-Atomistic Accuracy. Proteins. 2010; 78:1266–1281. [PubMed: 19967787]
47. Brooks BR, Brooks CL, Mackerell AD, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M. CHARMM: The Biomolecular Simulation Program. J Comput Chem. 2009; 30:1545–1614. [PubMed: 19444816]
48. Cheng YM, Gopal SM, Law SM, Feig M. Molecular Dynamics Trajectory Compression with a Coarse-Grained Model. IEEE Trans Comput Biol Bioinf. 2012; 6:476–486.
49. Predeus AV, Gul S, Gopal SM, Feig M. Conformational Sampling of Peptides in the Presence of Protein Crowders from AA/CG-Multiscale Simulations. J Phys Chem B. 2012; 116:8610–8620. [PubMed: 22429139]
50. Lee MS, Feig M, Salsbury FR, Brooks CL. New Analytic Approximation to the Standard Molecular Volume Definition and its Application to Generalized Born Calculations. J Comput Chem. 2003; 24:1348–1356. [PubMed: 12827676]
51. Wesson L, Eisenberg D. Atomic Solvation Parameters Applied to Molecular-Dynamics of Proteins in Solution. Protein Sci. 1992; 1:227–235. [PubMed: 1304905]
52. Lee MS, Salsbury FR Jr, Brooks CL III. Novel Generalized Born Methods. J Chem Phys. 2002; 116:10606–10614.
53. Kar P, Gopal SM, Cheng YM, Panahi A, Feig M. Transferring the PRIMO Coarse-Grained Force Field to the Membrane Environment: Simulation of Proteins and Helix-Helix Association. J Chem Theory Comput. 2014; 10:3459–3472. [PubMed: 25136271]

54. Tanizaki S, Feig M. Molecular Dynamics Simulations of Large Integral Membrane Proteins with an Implicit Membrane Model. J Phys Chem B. 2006; 110:548–556. [PubMed: 16471567]
55. Best RB, Zhu X, Shim J, Lopes P, Mittal J, Feig M, MacKerell AD Jr. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone φ, ψ and Side-Chain χ1 and χ2 Dihedral Angles. J Chem Theory Comput. 2012; 8:3257–3273. [PubMed: 23341755]
56. Best RB, Mittal J, Feig M, MacKerell AD. Inclusion of Many-Body Effects in the Additive CHARMM Protein CMAP Potential Results in Enhanced Cooperativity of α-Helix and β-Hairpin Formation. Biophys J. 2012; 103:1045–1051. [PubMed: 23009854]
57. Huang J, MacKerell AD. CHARMM36 All-Atom Additive Protein Force Field: Validation Based on comparison to NMR Data. J Comput Chem. 2013; 34:2135–2145. [PubMed: 23832629]
58. Arora K, Brooks CL. Large-Scale Allosteric Conformational Transitions of Adenylate Kinase Appear to Involve a Population-Shift Mechanism. Proc Natl Acad Sci US A. 2007; 104:18496–18501.
59. Vonrhein C, Schlauderer GJ, Schulz GE. Movie of the Structural-Changes during a Catalytic Cycle of Nucleoside Monophosphate Kinases. Structure. 1995; 3:483–490. [PubMed: 7663945]
60. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. J Comput Phys. 1977; 23:327–341.
61. Darden TA, York D, Pedersen LG. Particle-Mesh Ewald: An N log(N) Method for Ewald Sums in Large Systems. J Chem Phys. 1993; 98:10089–10092.
62. Jo S, Kim T, Iyer VG, Im W. CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. J Comput Chem. 2008; 29:1859–1865. [PubMed: 18351591]
63. Jana B, Adkar BV, Biswas R, Bagchi B. Dynamic Coupling Between the LID and NMP Domain Motions in the Catalytic Conversion of ATP and AMP to ADP by Adenylate Kinase. J Chem Phys. 2011; 134
64. Wang JN, Shao Q, Xu ZJ, Liu YT, Yang Z, Cossins BP, Jiang HL, Chen KX, Shi JY, Zhu WL. Exploring Transition Pathway and Free-Energy Profile of Large-Scale Protein Conformational Change by Combining Normal Mode Analysis and Umbrella Sampling Molecular Dynamics. J Phys Chem B. 2014; 118:134–143. [PubMed: 24350625]
65. Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JM. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules.1. The Method. J Comput Chem. 1992; 13:1011–1021.
66. Feig M, Karanicolas J, Brooks CL III. MMTSB Tool Set: Enhanced Sampling and Multiscale Modeling Methods for Applications in Structural Biology. J Mol Graph Modell. 2004; 22:377–395.
67. Beauchamp KA, Bowman GR, Lane TJ, Maibaum L, Haque IS, Pande VS. MSMBuilder2: Modeling Conformational Dynamics on the Picosecond to Millisecond Scale. J Chem Theory Comput. 2011; 7:3412–3419. [PubMed: 22125474]
68. Shen L, Hu H. Resolution-Adapted All-Atomic and Coarse-Grained Model for Biomolecular Simulations. J Chem Theory Comput. 2014; 10:2528–2536. [PubMed: 26580773]
69. Kabsch W, Sander C. Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features. Biopolymers. 1983; 22:2577–2637. [PubMed: 6667333]
70. MacKerell AD Jr, Feig M, Brooks CL III. Extending the Treatment of Backbone Energetics in Protein Force Fields: Limitations of Gas-Phase Quantum Mechanics in Reproducing Protein Conformational Distributions in Molecular Dynamics Simulations. J Comput Chem. 2004; 25:1400–1415. [PubMed: 15185334]
71. Song HD, Zhu FQ. Conformational Dynamics of a Ligand-Free Adenylate Kinase. Plos One. 2013; 8
72. MacKerell AD Jr, Feig M, Brooks CL III. Improved Treatment of the Protein Backbone in Empirical Force Fields. J Am Chem Soc. 2004; 126:698–699. [PubMed: 14733527]
73. Hanson JA, Duderstadt K, Watkins LP, Bhattacharyya S, Brokaw J, Chu JW, Yang H. Illuminating the Mechanistic Roles of Enzyme Conformational Dynamics. Proc Natl Acad Sci US A. 2007; 104:18055–18060.

74. Aden J, Wolf-Watz M. NMR Identification of Transient Complexes Critical to Adenylate Kinase Catalysis. J Am Chem Soc. 2007; 129:14003–14012. [PubMed: 17935333]

75. Shapiro YE, Meirovitch E. Activation Energy of Catalysis-Related Domain Motion in *E. coli* Adenylate Kinase. J Phys Chem B. 2006; 110:11519–11524. [PubMed: 16771428]

76. Matsunaga Y, Fujisaki H, Terada T, Furuta T, Moritsugu K, Kidera A. Minimum Free Energy Path of Ligand-Induced Transition in Adenylate Kinase. Plos Comp Biol. 2012; 8:e1002555.

77. Weinan E, Ren WQ, Vanden-Eijnden E. Finite Temperature String Method for the Study of Rare Events. J Phys Chem B. 2005; 109:6688–6693. [PubMed: 16851751]

78. Adkar BV, Jana B, Bagchi B. Role of Water in the Enzymatic Catalysis: Study of ATP plus AMP -> 2ADP Conversion by Adenylate Kinase. J Phys Chem A. 2011; 115:3691–3697. [PubMed: 20836529]

79. Lou HF, Cukier RI. Molecular Dynamics of apo-Adenylate Kinase: A Distance Replica Exchange Method for the Free Energy of Conformational Fluctuations. J Phys Chem B. 2006; 110:24121–24137. [PubMed: 17125384]

80. Ensing B, Nielsen SO, Moore PB, Klein ML, Parrinello M. Energy Conservation in Adaptive Hybrid Atomistic/Coarse-Grain Molecular Dynamics. J Chem Theory Comput. 2007; 3:1100–1105. [PubMed: 26627429]

81. Feig M, Chocholousova J, Tanizaki S. Extending the Horizon: Towards the Efficient Modeling of Large Biomolecular Complexes in Atomic Detail. Theor Chem Acc. 2006; 116:194–205.

82. Jung J, Mori T, Kobayashi C, Matsunaga Y, Yoda T, Feig M, Sugita Y. GENESIS: A Hybrid-Parallel and Multi-Scale Molecular Dynamics Simulator with Enhanced Sampling Algorithms for Biomolecular and Cellular Simulations. Wiley Interdiscip Rev: Comput Mol Sci. 2015; 5:310–323. [PubMed: 26753008]

83. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. Scalable Molecular Dynamics with NAMD. J Comput Chem. 2005; 26:1781–1802. [PubMed: 16222654]

84. Salomon-Ferrer R, Götz AW, Poole D, Le Grand S, Walker RC. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. J Chem Theory Comput. 2013; 9:3878–3888. [PubMed: 26592383]

85. Friedrichs MS, Eastman P, Vaidyanathan V, Houston M, Legrand S, Beberg AL, Ensign DL, Bruns CM, Pande VS. Accelerating Molecular Dynamic Simulation on Graphics Processing Units. J Comput Chem. 2009; 30:864–872. [PubMed: 19191337]

86. Jung J, Naurse A, Kobayashi C, Sugita Y. Graphics Processing Unit Acceleration and Parallelization of GENESIS for Large-Scale Molecular Dynamics Simulations. J Chem Theory Comput. 2016; 12:4947–4958. [PubMed: 27631425]

87. Arthur EJ, Brooks CL. Parallelization and Improvements of the Generalized Born Model with a Simple SWitching Function for Modern Graphics Processors. J Comput Chem. 2016; 37:927–939. [PubMed: 26786647]

88. Panahi A, Feig M. Dynamic Heterogeneous Dielectric Generalized Born (DHDGB): An Implicit Membrane Model with a Dynamically Varying Bilayer Thickness. J Chem Theory Comput. 2013; 9:1709–1719. [PubMed: 23585740]

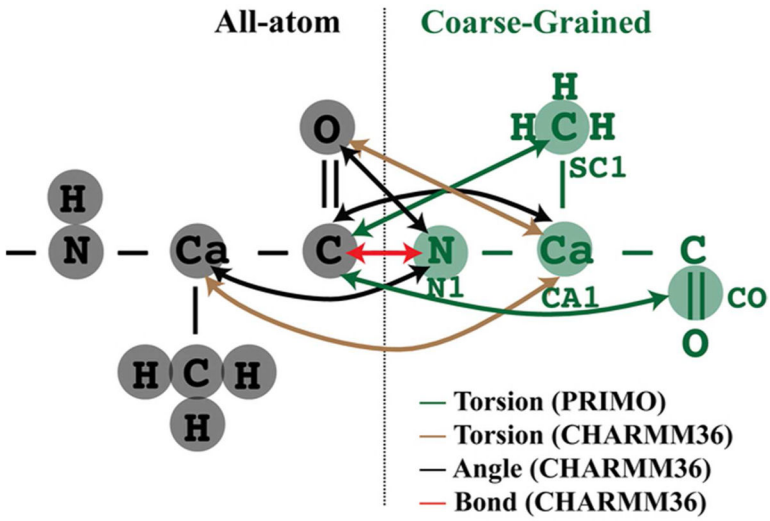


Figure 1.
Hybrid AA/CG coupling scheme.

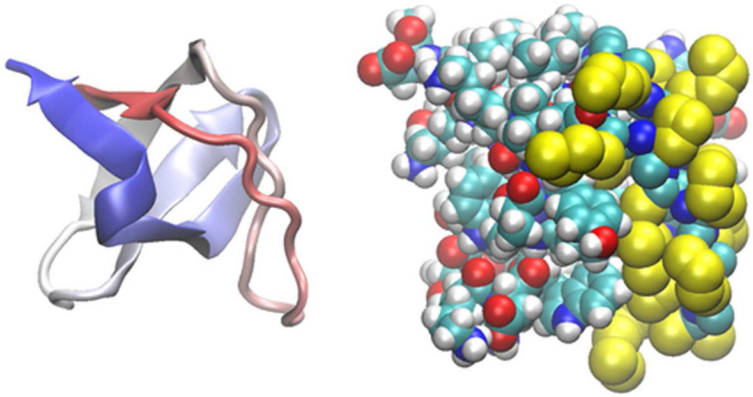


Figure 2.

SH3 domain (PDB code: 1SHG) in cartoon representation (left, colored by index from red to blue) and oriented in the same way at the AA/CG level (right) with residues 6–12 and 30–62 at the atomic level and residues 13–29 described by PRIMO CG model. Atoms are colored by atom type: cyan (C); white (H); red (O); blue (N); yellow (CG side chain).

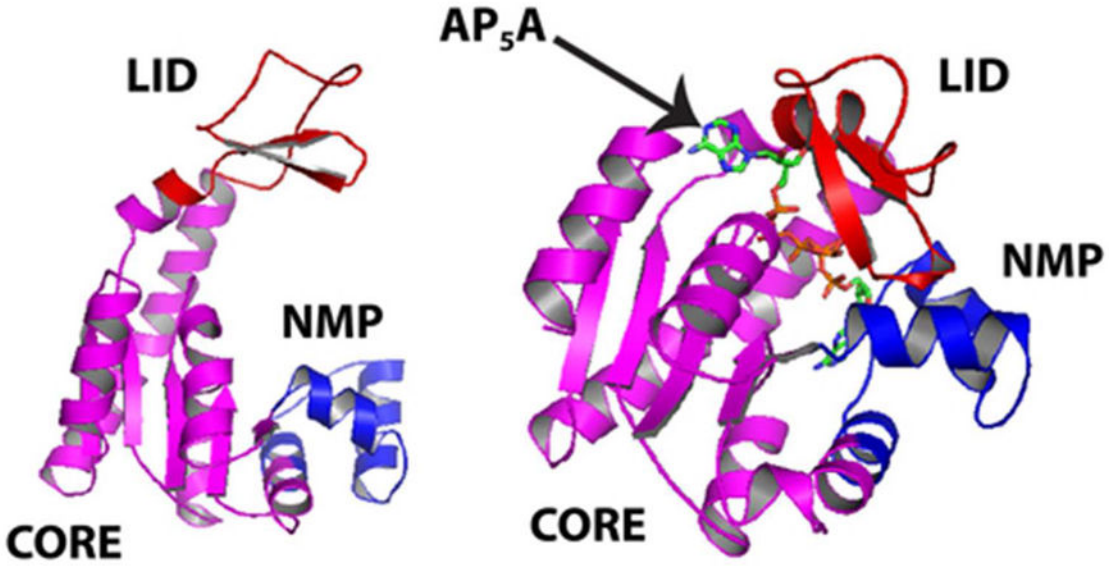


Figure 3.

The open (left) and closed (right) forms of adenylate kinase. The inhibitor P1,P5-Di(adenosine-5′)pentaphosphate (AP_5A) present in the crystal structure is shown in stick representation to indicate where the nucleotides would bind but no ligands were present in the simulations reported in this study.

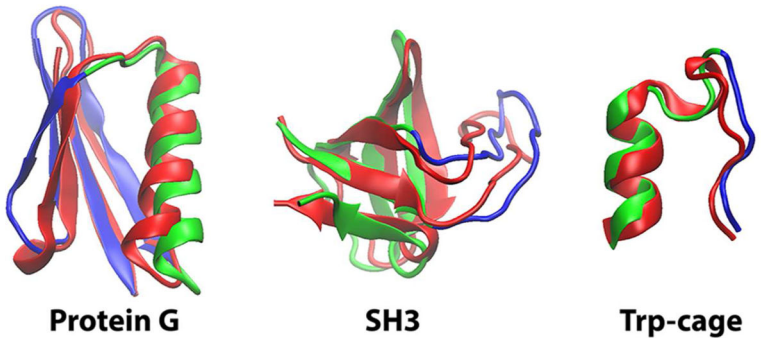


Figure 4.
Ensemble-averaged structures from the AA/CG simulations (green: AA region, blue: CG region) compared to experimental structures (red).

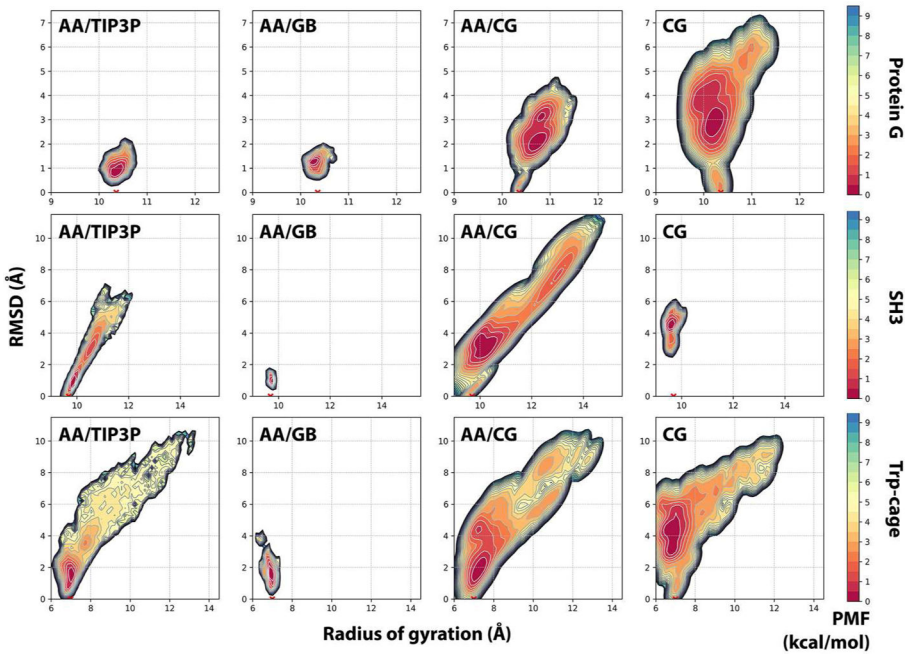


Figure 5.

Potentials of mean force as a function of Cα-based radius of gyration and RMSD with respect to experimental structures for protein G (top), SH3 domain (middle), and Trp-cage (bottom) with with fully atomistic simulations with explicit solvent (AA/TIP3P), AA/GBMV (AA/GB), the hybrid AA/CG model (AA/CG), or the PRIMO force field (CG).

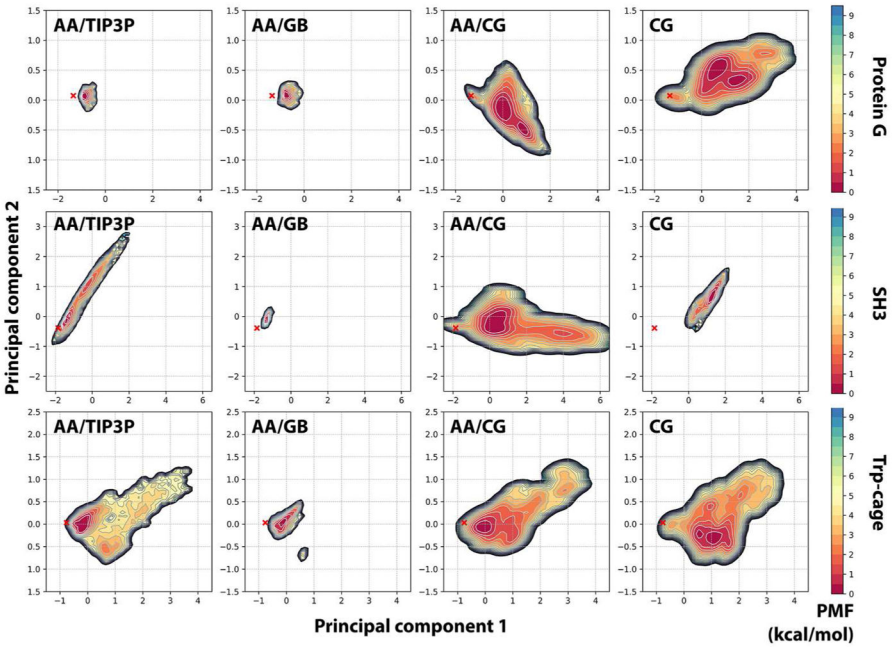


Figure 6.

Potentials of mean force as a function of principal coordinates for protein G (top), SH3 domain (middle), and Trp-cage (bottom) with with fully atomistic simulations with explicit solvent (AA/TIP3P), AA/GBMV (AA/GB), the hybrid AA/CG model (AA/CG), or the PRIMO force field (CG). For each system, common principal coordinates were determined from combining all of the trajectories.

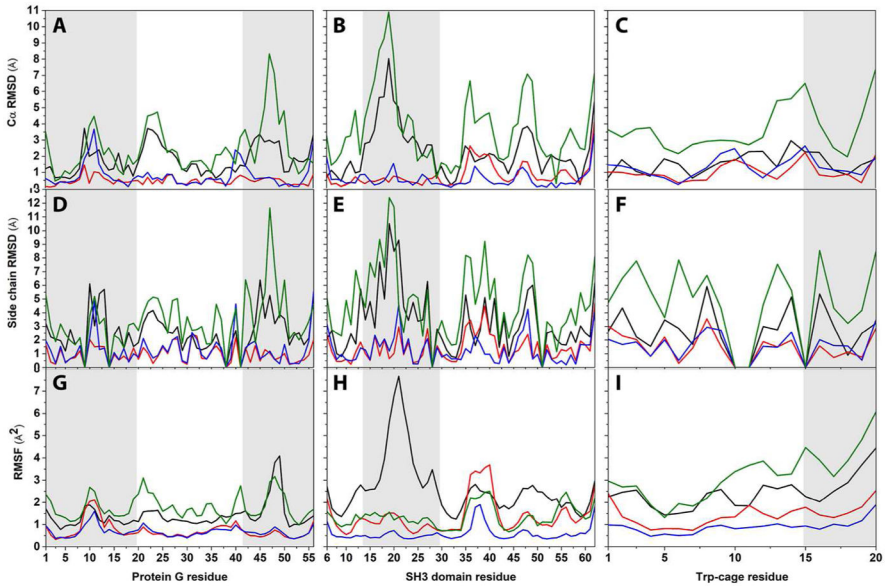


Figure 7.

Per-residue root mean square deviations for C_α coordinates (top row, A–C) and side chains (middle row, D–F) and root mean square fluctuations for C_α coordinates (bottom row, G–I) with respect to the experimental structures for protein G (PDB code: 3GB1, left column, A/D/G), the SH3 domain (PDB code: 1SHG, center, B/E/H), and Trp-cage (PDB code: 1L2Y, right, C/F/I) with the hybrid AA/CG model (AA/CG, black), fully atomistic simulations with explicit solvent (AA/TIP3P, red), AA/GBMV (AA/GB, blue), or coarse-grained simulations using the PRIMO force field (CG, green). Shaded areas indicate regions modeled at the CG level in the AA/CG hybrid model.

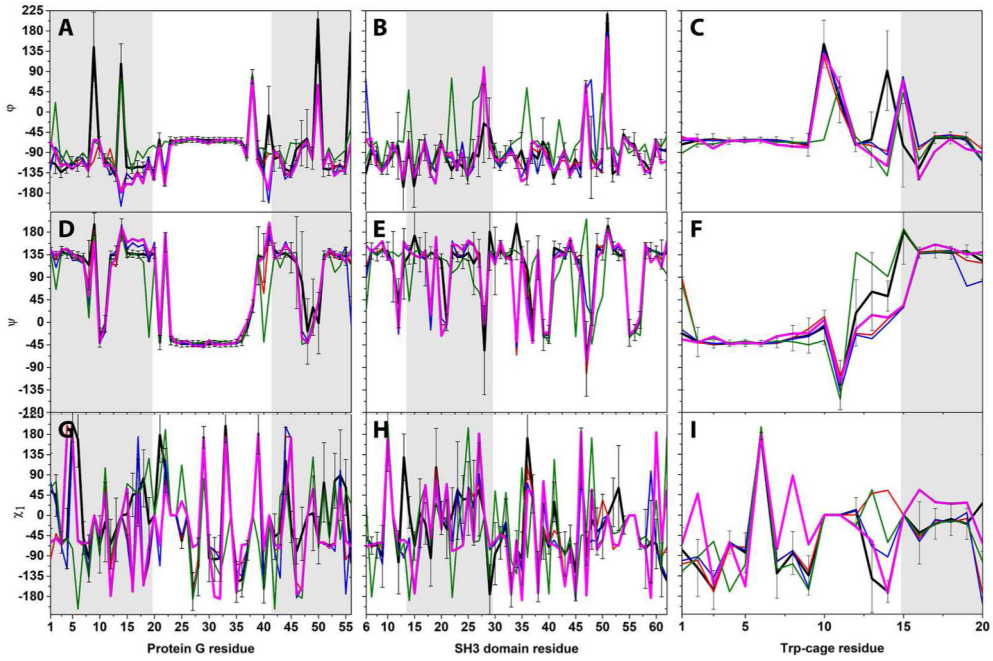


Figure 8.

Per-residue average backbone torsions φ (top row, A–C) and ψ (middle row, D–F) and side chain χ_1 torsion (bottom row, G–I) for protein G (left column, A/D/G), the SH3 domain (center, B/E/H), and Trp-cage (right, C/F/I) with the hybrid AA/CG model (AA/CG, black), fully atomistic simulations with explicit solvent (AA/TIP3P, red), AA/GBMV (AA/GB, blue), or coarse-grained simulations using the PRIMO force field (CG, green). Torsion angles in the experimental structures are shown in magenta. Shaded areas indicate regions modeled at the CG level in the AA/CG hybrid model.

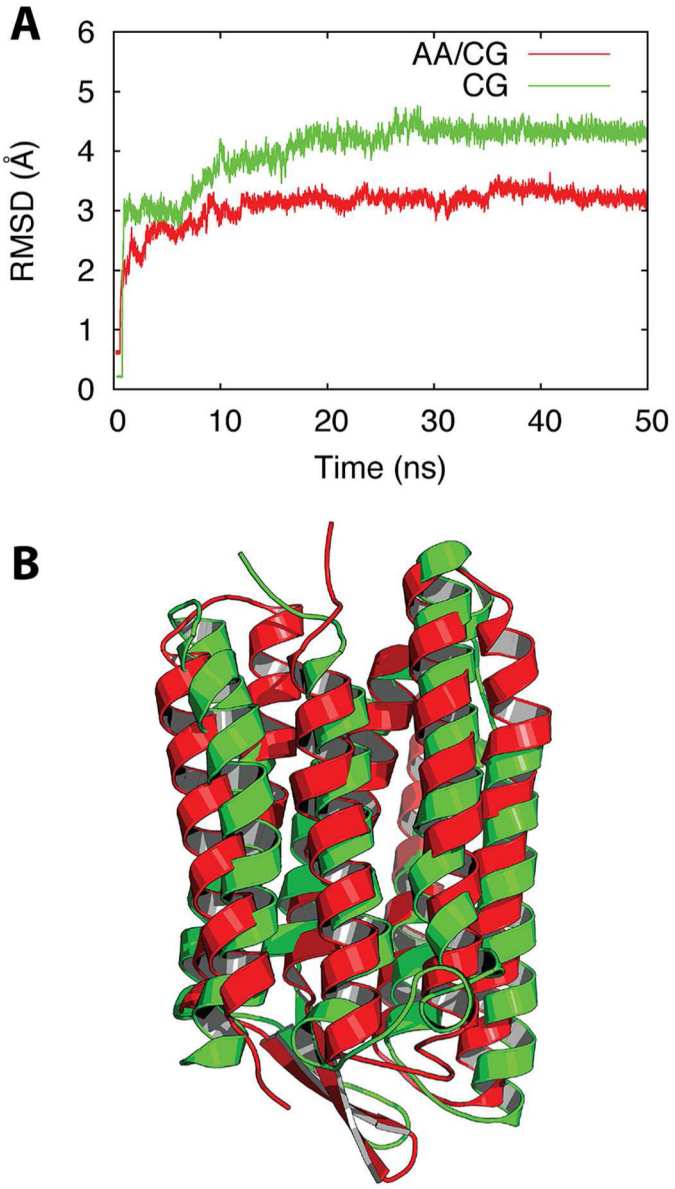


Figure 9.

Time evolution of root-mean-squared-deviation (RMSD) of C_α atoms with respect to the experimental structure for bacteriorhodopsin (PDB code: 1QHJ). Results obtained from the hybrid AA/CG simulation (red) are shown as well as coarse-grained simulations using the PRIMO-M force field (green).

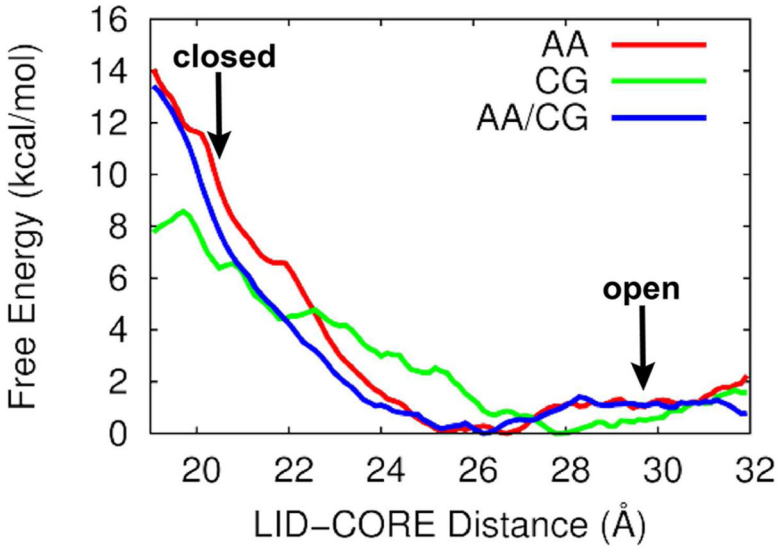


Figure 10.

Free energy profiles of the LID-CORE motion in adenylate kinase obtained from AA/CG (blue) umbrella sampling, compared with results from atomistic (red) and coarse-grained simulations (green).

Table 1

Overview of the systems studied here at AA, CG, and AA/CG levels.
AA/exp refers to either AA/TIP3P for the soluble systems or an explicit POPC lipid bilayer surrounded by explicit water for bacteriorhodopsin.

	Trp-cage	SH3	Protein G	Bacteriorhodopsin	Adenylate kinase
AA/TIP3P					
# atoms	7898	10845	16090	32344	55499
AA res.	1–20	6–62	1–56	5–232	1–214
AA/GBMV					
# atoms	313	964	864	3599	3350
AA res.	1–20	6–62	1–56	5–232	1–214
AA/CG					
# atoms	252	782	496	2245	1833
AA res.	1–14	6–12; 30–62	21–41	5–103	30–67; 118–160
CG res.	15–20	13–29	1–20; 42–56	104–232	1–29; 68–117; 161–214
CG					
# atoms	105	333	310	1207	1165
CG res.	1–20	6–62	1–56	5–232	1–214

Table 2

C_α RMSD from experimental structures and radius of gyration.

	Model	Trp-Cage	SH3	Protein G
Average C_α RMSD (Å)	AA/CG	2.66 *(0.13)*	3.71 *(0.41)*	2.42 *(0.16)*
	CG	4.28 *(0.13)*	4.44 *(0.10)*	3.29 *(0.07)*
	AA/GBMV	1.66 *(0.004)*	1.01 *(0.05)*	1.25 *(0.01)*
	AA/TIP3P	1.68 *(0.08)*	1.47 *(0.50)*	0.95 *(0.02)*
C_α RMSD of average structure (Å)	AA/CG	1.70	2.75	2.00
	CG	4.03	4.35	2.94
	AA/GBMV	1.49	0.79	1.05
	AA/TIP3P	1.19	1.09	0.55
Average C_α radius of gyration R_g (Å)	AA/CG	7.45 *(0.07)*	10.50 *(0.30)*	10.74 *(0.02)*
	CG	6.97 *(0.15)*	9.61 *(0.04)*	10.19 *(0.03)*
	AA/GBMV	6.94 *(0.001)*	9.74 *(0.01)*	10.29 *(0.01)*
	AA/TIP3P	7.04 *(0.04)*	10.02 *(0.16)*	10.37 *(0.03)*
	Exp. (from PDB)	7.00	9.63	10.36

Standard errors are provided in parentheses.

Table 3

Average solvent-accessible surface area (SASA) and molecular volume obtained from AA/CG, CG, and AA simulations in comparison with SASA and molecular volume of PDB structures. AA SASA and CG SASA refer to the surface areas of the subset of residues simulated at AA or CG levels in the AA/CG model (see Table 1) in the context of the entire structure. Standard errors are provided in the parentheses.

	Model	**Total SASA (Å^2)**	**AA SASA (Å^2)**	**CG SASA (Å^2)**	**Volume (Å^3)**
Protein G	AA/CG	4455.7 *(10.9)*	1591.3 *(7.3)*	2864.4 *(16.1)*	6746.9 *(1.1)*
	AA/TIP3P	3921.5 *(16.5)*	1430.9 *(6.5)*	2490.6 *(11.3)*	6769.4 *(0.8)*
	AA/GBMV	3854.2 *(11.3)*	1421.5 *(0.3)*	2432.7 *(11.6)*	6759.8 *(1.5)*
	CG	4318.3 *(12.1)*	1444.2 *(18.0)*	2874.1 *(17.0)*	6698.1 *(2.3)*
	PDB	3906.8	1423.9	2483.0	6772.1
SH3	AA/CG	4918.5 *(66.7)*	3246.8 *(36.3)*	1671.6 *(39.2)*	7462.5 *(3.1)*
	AA/TIP3P	4357.3 *(59.4)*	2993.1 *(48.9)*	1364.2 *(10.9)*	7458.3 *(3.6)*
	AA/GBMV	4273.4 *(12.1)*	2906.6 *(5.2)*	1366.8 *(6.9)*	7438.3 *(0.7)*
	CG	4582.2 *(34.2)*	3091.0 *(28.4)*	1491.2 *(5.8)*	7325.8 *(6.9)*
	PDB	3990.1	2720.7	1269.4	7425.7
Trp-cage	AA/CG	2203.0 (15.5)	1476.8 (11.6)	726.2 (4.7)	2483.4 *(1.9)*
	AA/TIP3P	2002.5 *(7.4)*	1337.4 *(5.2)*	665.1 *(2.7)*	2467.4 *(0.3)*
	AA/GBMV	1999.4 *(1.8)*	1373.2 *(1.8)*	626.2 *(1.5)*	2467.6 *(0.1)*
	CG	2204.7 *(23.2)*	1525.3 *(15.0)*	679.4 *(11.8)*	2466.3 *(1.7)*
	PDB	1926.4	1266.4	660.1	2468.1

Table 4

CPU time spent for 10 ps of MD simulation (5,000 steps with 2 fs time step for AA/exp, AA/GBMV, and AA/CG models and 2,500 steps with 4 fs time step for CG model) for five different protein systems. AA/exp refers to simulations with explicit water or water and POPC lipids (bacteriorhodopsin). All the simulations were performed in serial on a single core of an Intel E5–2680v3 processor (2.5 GHz) using CHARMM c41a1. CPU times are provided in seconds. Speedups with respect to AA/GBMV are provided in parentheses.

System	AA/exp.	AA/GBMV	AA/CG	CG
Trp-cage	802	216	150 *(1.4)*	21 *(10.3)*
SH3	1139	820	584 *(1.4)*	80 *(10.3)*
Protein G	1689	715	316 *(2.3)*	74 *(9.7)*
Bacteriorhodopsin	3090	3219	1786 *(1.8)*	307 *(10.5)*
Adenylate kinase	3101	2915	1374 *(2.1)*	293 *(9.9)*