Recent Advances in Transferable Coarse-Grained Modeling of Proteins

Parimal Kar[1] and **Michael Feig**[1,2,*]

[1]Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824, USA

[2]Department of Chemistry, Michigan State University, East Lansing, MI 48824, USA

Abstract

Computer simulations are indispensable tools for studying the structure and dynamics of biological macromolecules. Biochemical processes occur on different scales of length and time. Atomistic simulations cannot cover the relevant spatiotemporal scales at which the cellular processes occur. To address this challenge, coarse-grained (CG) modeling of the biological systems are employed. Over the last few years, many CG models for proteins continue to be developed. However, many of them are not transferable with respect to different systems and different environments. In this review, we discuss those CG protein models that are transferable and that retain chemical specificity. We restrict ourselves to CG models of soluble proteins only. We also briefly review recent progress made in the multi-scale hybrid all-atom/coarse-grained simulations of proteins.

Keywords

coarse-graining; force field; multi-scale; molecular dynamics; protein folding; protein-protein interactions

1. Introduction

Classical all-atom molecular dynamics simulations with explicit solvent and ions are routinely used to investigate the dynamics of biological macromolecules. Such simulations can reveal the mechanism of biochemical processes occurring on different scales of length and time[1]. With the availability of highly optimized atomistic force fields, specialized supercomputers, scalable codes, and enhanced sampling methodologies, one can study the dynamics of biomolecules in a realistic fashion and explain and predict experimental results. Most notable recent highlights are folding studies of small proteins by Schulten et al.[2] and Shaw et al.[3] The Schulten group has obtained a continuous trajectory of 10 μs while the Shaw group has performed 1 ms long simulation that would have been unthinkable not long ago. Other examples include simulations of proteins in cellular environments[4–6] to investigate the effect of cellular-like environments on the stability and diffusive properties of

---

[*]Corresponding author: E-mail: feig@msu.edu, Phone: 517-432-7439, Fax: 517-353-9334.

biological macromolecules and simulations of very large complexes such as RNA polymerase II[7] and MutSα/MutSβ[8,9]. Despite such successes, atomistic simulations are still several magnitudes away from being able to effectively cover the spatiotemporal scales of most cellular processes.

To address this challenge, the most promising strategy is a reductionist approach, i.e. a simplification of the underlying model that is used to describe the biological system (see Figure 1). This can be achieved by coarse-graining of the system. In coarse-grained (CG) models several atoms are grouped into a single particle (CG bead) and the interactions between these CG beads are then described by effective potentials that incorporate both energetic and entropic contributions resulting from integrating out (or averaging over) the neglected atomistic details. The coarse-graining of the system can greatly reduce the number of degrees of freedom of the system compared to a complete all-atom description of the system. As a result, less computational power is required to simulate the system. Furthermore, the interactions between CG sites are usually "softer" which allows for longer integration time steps in molecular dynamics simulations. Typical time steps that can be used with CG models are tens of femtoseconds compared to 1–2 fs used in all-atom MD simulations. CG models combined with the dissipative particle dynamics (DPD) allow time steps of 100 femtosecond or above. Interactions in CG models tend to be of shorter range than in all-atom models due to the use of effective energy terms and typically the expensive treatment of long-range electrostatics can be avoided. This results in additional significant speedups in CG simulations compared with atomistic simulations. Finally, CG simulations generate inherently accelerated dynamics since the energy landscapes in CG models are smoother compared with atomistic models. Because of these advantages over fully atomistic simulations, CG models have the potential to cover significantly larger spatial and longer temporal scales compared to all-atom models.

In the past few years there has been a flurry of CG models for proteins with varying complexities[10,11]. These CG models are employed for various applications, such as study of protein folding[12,13] problems, protein–protein docking[14], structure predictions[15,16], scoring protein decoys[17], structural features of viral capsid[18,19], structure and dynamics of membrane or protein-membrane systems[20], and influence of certain amino acid substitutions in the HIV-1 Gag polyprotein and how the resulting physical effects influence HIV-1 virion formation and structure[21].

There is no unique way of coarse-graining a protein[22]. It depends on the levels of granularity and specific applications. The CG mapping schemes of proteins fall into two major categories: shape-based[19] and residue-based approaches[11]. The shape-based CG models emphasize outwards appearance and are thus not suitable for studying protein folding problems or distinguishing the dynamics between the wild-type and mutant variants of a protein. Residues-based models are suited best for studying internal dynamics and effects of chemical specificity of different residues.

Depending on the CG mapping scheme and specific applications, CG force fields may range from simple effective interaction potentials to more realistic all-atom like force fields. Generally, two different approaches, namely bottom-up (structure-based coarse-graining)

and top-down (thermodynamic-based coarse-graining), are followed to generate a CG force field. In the bottom-up approach, CG force fields are parameterized based on comparisons to reference atomistic simulations. Several approaches, such as inverse Monte Carlo (IMC)[23], iterative Boltzmann inversion (IBI)[24], force matching (FM)[25–27], fluctuation matching[28], minimization of relative entropy[29], or conditional reversible work (CRW)[30,31] are employed in bottom-up approaches to match a set of target distributions from atomistic simulations. The bottom-up approaches are also highly amenable to multi-scale approaches since the fine grained (AA) information is propagated to a lower resolution (CG) model. On the other hand, in top-down approaches, CG force fields may be parameterized against experimental data, in particular thermodynamic data such as oil/water partitioning coefficients and density. Among various bottom-up approaches, IBI and FM methods are most popular and are frequently used to parameterize CG force fields for biomolecular simulations. IBI reproduces the structure of the underlying atomistic model by means of an iterative optimization scheme through which the effective potentials between CG sites are obtained. This method is motivated by Henderson's uniqueness theorem[32] which states that, for a given pair radial distribution function, there is a unique underlying pair potential that will produce it. On the other hand, FM does not depend on the pair correlation functions. Instead, the idea of FM is to match the average force on CG beads in the CG system to that expected from the atomistic model. We are not going to discuss the underlying theories of different parameterization procedures employed in the bottom-up approaches and for an extended overview of such topics, the reader is referred to a recent review by Noid[33].

A major challenge remains in developing CG models that are highly predictive with respect to (unknown) specific behavior of specific systems rather than simply recapitulating either results from atomistic simulations or the generally expected characteristics of protein biophysics. This generally requires that CG models exhibit a sufficient degree of both transferability and chemical specificity. The range of CG models is very broad given different levels of coarse-graining and various philosophies of the parameterization of the CG force fields. Here, we will focus primarily on recent advances in developing such transferable CG force fields as they show the most promise in broad future applications of CG methodologies.

This review is organized as follows: First, we give a review of popular CG models for proteins. Next, we discuss recent advances in hybrid all-atom/coarse-grained (AA/CG) simulation methodologies. Finally, we conclude with a short outlook.

2. CG Models of Protein

Generally, a CG model involves the following four choices[34]: (i) resolution of the model appropriate for the size of the system and the types of questions that are going to be addressed; (ii) mapping of the CG beads to AA sites; (iii) type of interaction potential, and (iv) parameterization of the interaction potential to faithfully describe the energetics and dynamics of the biological system in comparison with experiment and AA simulations. CG models of varying complexity have been proposed for proteins since the pioneering work of Levitt and Warshel[35]. Their resolution may vary from one or a few beads per amino acids to near atomistic resolutions[11,36–43]. The coarser the description, the larger the gain in

computational speedup. However, coarser models lack transferability and may not be suitable for multi-scale simulations.

Early on, one-bead models[38,44,45] were most popular. For instance, Tozzini and McCammon have developed a Cα based single bead CG model to study the dynamics of flap opening in HIV-1 protease[38]. In their model, each amino acid is represented by a single bead placed on the Cα atom. The force field was parameterized based on the Boltzmann inversion procedure. A similar approach was adopted by Korkuta and Hendrickson[45] to build the virtual atom molecular mechanics (VAMM) force field. To study the disordered state of proteins, Ghavami and co-workers[42] have proposed a one-bead-per-amino-acid model. Generally, one-bead models are highly system-specific, such as the widely used G̅ models, and they typically lack chemical specificity of amino acid side-chains. Furthermore, one-bead models are typically not well suited for multi-scale simulations because of their low resolution.

In contrast to other CG models, a model developed by the Voth group did not rely on a predefined potential but instead utilized the multi-scale coarse-graining (MS-CG) method[25,46,47]. In this approach, the CG model and interaction potential was derived via a force-matching method in which atomistic force information from reference simulations is matched to forces from CG potentials within a variational framework[26,27]. One of the advantages of the MS-CG approach is that it can be applied at arbitrary resolutions. The MS-CG method has been applied to the study of protein folding and dynamics[13,48]. However, MS-CG potentials are commonly not transferable. These models will not be described here. The reader is referred to earlier reviews on such models[48–51].

Side-chain chemical specificity is essential for transferability and is the key for describing packing interactions within and between proteins. Adding one or several beads for the side-chain in addition to a single bead for the backbone can greatly improve the specificity of local interactions[43,52]. A detailed description of the backbone, on the other hand, allows a faithful reproduction of secondary structure propensities. In the last few years, a number of CG models were developed that have relatively high resolution, with several beads per residue, and that take into account the chemical specificity of side-chains. These models employ different types of interaction potentials but generally offer at least some degree of transferability. In the following, the main features of the most prominent examples of such models are described. Table 1 summarizes mapping schemes and different force field terms for these models.

UNRES Model

The UNited RESidue (UNRES) model was initially developed by the Scheraga group almost two decades ago as a knowledge-based CG model[53–55]. However, over the years, it has undergone several refinement stages and finally evolved into a fully physics-based model. In the UNRES model[53–56], a polypeptide chain is represented as a sequence of α-carbon atoms with attached united side-chains (SC), whose sizes depend on the nature of the amino acid residue and united peptide groups (P), each of the latter being positioned in the middle between two consecutive Cα atoms (Figure 2). The α-carbon atoms are linked by virtual Cα…Cα bonds of length 3.8 Å corresponding to *trans* peptide groups. It should be noted

here that these Cα atoms are not interaction sites *per se* but rather they assist only in the definition of the geometry. Each SC is represented as an ellipsoid and is attached to the respective Cα atom by a virtual Cα…SC bond.

The force field is derived as a restricted free energy (RFE) function, which corresponds to averaging the energy over the degrees of freedom that are neglected in the UNRES model. The UNRES effective energy function is given by equation 1[57].

$$
\begin{aligned}
U^{\mathrm{UNRES}} = {}& w_{\mathrm{SC}} \sum_{i<j} U_{\mathrm{SC_i SC_j}} + w_{\mathrm{SC_p}} \sum_{i \neq j} U_{\mathrm{SC_i P_j}} + \\
& w_{\mathrm{pp}} f_2(T) \sum_{i<j-1} U_{\mathrm{p_i p_j}} + w_{\mathrm{tor}} f_2(T) \sum_i U_{\mathrm{tor}}(\gamma_i) + \\
& w_{\mathrm{tord}} f_3(T) \sum_i U_{\mathrm{tord}}(\gamma_i, \gamma_{i+1}) + w_{\mathrm{b}} \sum_i U_{\mathrm{b}} + (\theta_i) + \\
& w_{\mathrm{rot}} \sum_i U_{\mathrm{rot}}(\alpha_{\mathrm{SC_i}}, \beta_{\mathrm{SC_i}}, \theta_i) + w_{\mathrm{bond}} \sum_i U_{\mathrm{bond}}(d_i) + \\
& \sum_{m=3}^{6} w_{\mathrm{corr}}^{m} f_m(T) U_{\mathrm{corr}}^{(m)} + w_{\mathrm{ss}} \sum_i U_{\mathrm{ss,i}}
\end{aligned}
\tag{1}
$$

where *w* represents the relative weights of each term. The initial three terms represent non-bonded interaction energies between SC-SC, SC-P, P-P beads. The next five terms describe the bonded interaction energies corresponding to torsion, double-torsion, angle, SC rotamers, and bonds, respectively. The last two remaining terms model multi-body (correlation) interaction energies and the energy of forming disulfide bonds, respectively. Notably, the correlation term is temperature dependent. The force field is optimized based on the concept of a hierarchical protein-energy landscape[58].

The potential function of UNRES is parameterized against free energies from atomistic polypeptide simulations in explicit water using a functional form that resembles atomistic force fields but with additional terms. In UNRES, solvent effects are represented implicitly. Applications of UNRES to date have been largely limited to *de novo* protein folding studies[12,59,60] and protein structure prediction[61,62]. Recently, Solernou and Fernandez-Recio[63] have developed a CG potential by extending the CG potential UNRES with additional terms accounting for a solvation energy contribution and an explicit electrostatic interactions to predict the protein-protein interactions.

Among all of the physics-based CG models, UNRES performs exceptionally well in folding proteins to their native-like conformations. This model has been used successfully in CASP structure prediction competition. For instance, the target T0215 was fold to a Cα RMSD of 3.5 Å from the native state. UNRES has been used to fold multi-chain proteins also. For example, UNRES was used to ab initio folding of the multi-chain protein with PDB ID 1G6U and the lowest RMSD obtained was 2.4 Å from the native state[64]. Their simulations reveal that individual chain folded independently to their native structure and then assembled into a complex structure.

The UNRES model is physically accurate and highly transferable but it suffers from a number of drawbacks. First, the model requires a very complex reverse-mapping scheme to

convert from UNRES to all-atom models and is thus difficult to use within the context of multi-scale AA/CG simulations. Second, UNRES is neither transferable to the lipid environments nor is it easy to extend the model for other biomolecules such as nucleic acids because the necessary model re-parameterization would likely be very time-consuming.

MARTINI Model

The most popular and most widely used CG model for lipids as well as proteins is MARTINI[65–67]. This model has been employed in a wide range of applications[68–76]. The model was initially developed for lipid molecules and later was extended to proteins and several other biomolecules. The MARTINI model is based on a four-to-one mapping, i.e., on average four heavy atoms plus associated hydrogens are represented by a single CG site (Figure 2). However, ring-like molecules (e.g., benzene, cholesterol, and several of the amino acids) are mapped with a higher resolution of up to two non-hydrogen atoms to one MARTINI particle. The backbone of each amino acid is modeled with a single bead placed at the center of mass of the backbone, while up to four beads are used for SC. In contrast to UNRES, solvent is treated at the CG level in MARTINI. Four real water molecules are mapped to a single CG water bead. Recently, polarizable CG water models[77,78] were also developed to be used with the MARTINI model. Ions are represented by a single CG bead, which represents both the ion itself and its first hydration shell. Based on the nature of different chemical groups, four main types of particle are assigned in the MARTINI model, namely polar (P), non-polar (N), apolar (C), and charged (Q). Within a main type, subtypes are assigned according to hydrogen bonding capabilities or the degree of polarity. In total, MARTINI has 18 particle types.

The MARTINI CG force field consists of generic bonded (bond, angle, dihedral, and improper dihedral) and non-bonded (van der Waals and electrostatics) interaction potentials. Notably, the electrostatic interaction is described by the Coulomb potential with a relative dielectric constant ($\varepsilon_{rel} = 15$) for explicit screening. However, a lower relative dielectric constant of 2.5 is suggested when combining MARTINI with a polarizable water model. The parameterization of this model rely on a combination of bottom-up and top-down approaches. For instance, a bottom-up approach was adopted to derive the bonded terms and were obtained from the underlying atomistic geometry or by comparing to atomistic simulations. On the other hand, the non-bonded interaction potentials were optimized based on a systematic comparison to experimental theromodynamic quantities, such as free energy of hydration and vaporization, and the partitioning free energies between oil and aqueous phases. Recently, the non-bonded parameters were further optimized[79] based the binding studies of Wimley-White pentapeptides to a 2-oleoyl-1-pamlitoyl-*sn*-glyecro-3-phosphocholine (POPC) bilayer[80] and SC analog dimerization free energies[81]. This led to the addition of an explicit polarization term to polar SCs such as glutamine and asparagine. The MARTINI CG model can achieve about a factor of 75–100 speedup with respect to fully atomistic simulations[82].

MARTINI suffers from not being able to accurately represent amino-acid specific secondary structure propensities. As a consequence, a weak bias has to be employed to maintain secondary structures based on what is known from native structures. Therefore, MARTINI is

not suitable for modeling processes in which secondary structure changes and, more generally, folding and unfolding play a substantial role. However, movements of secondary structures with respect to each other can be studied with the MARTINI model. Furthermore, although MARTINI is very successful in modeling interactions with lipid bilayers and lipid-water interfaces, it is not applicable to systems involving solid/fluid or gas/fluid interfaces because of the fluid-phase based parameterization.

It is worth mentioning here that Peraro and coworkers[83] have adopted the MARTINI mapping scheme and have developed an electrostatic-consistent CG potential for proteins in which they have introduced an explicit dipole defined by three consecutive Cα beads along with the treatment of non-radial dipole-dipole interactions in dynamics. As a result, in contrast to MARTINI, this model can maintain stable secondary elements of unspecific proteins and sample conformational transitions.

OPEP Model

The OPEP (Optimized Potential for Efficient structure Prediction)[16,84–86] is a transferable CG force field for proteins developed by Derreumaux and coworkers. OPEP is an intermediate resolution model where each amino acid is represented with six beads. The backbone is retained in full atomic detail while SCs are represented by a single bead or centroid (Figure 2). However, the amino acid proline (Pro) is represented by all heavy atoms. The centroids are placed at the center of mass of the heavy atoms and calculated employing the rotamer distribution of each SC. Each SC bead is described by an appropriate van der Waals radius, and their positions are defined with respect to the backbone heavy atoms N, C_{α}, and C'. The OPEP energy function consists of bonded and non-bonded terms. Bonded terms include bonds, angles, dihedrals, and improper torsions (to maintain the *cis* or *trans* conformation of the peptide bond and the chirality of the amino acid). Additional terms confine the dihedrals to the most populated regions in the Ramachandran plot via harmonic barriers. The force constants and equilibrium values associated with these terms were derived from the AMBER all-atom force field[87]. Non-bonded interactions in OPEP consist of van der Waals interactions and explicit hydrogen bonding potentials. The hydrogen-bonding potential consists of two-body and four-body terms, and the later term was used to capture the strong cooperative nature of the amide hydrogen bonds. The van der Waals interactions are knowledge-based as van der Waals radii were calculated using a dataset of thousands of Protein Data Bank (PDB) structures. The van der Waals interactions combine backbone-backbone, backbone-SC terms with 210 SC pair interactions. The aqueous solvent is treated implicitly in OPEP. A genetic algorithm using a scoring function, which requires that the native structure has the lowest energy and that native-like structures have an energy higher than the native structure but lower than non-native conformations, was used to optimize the OPEP force field. In OPEPv4, the van der Waals term was composed of an r^{-8} term combined with an analytical formulation that quickly vanishes as a function of distance[88]. This analytical function has a steeper profile with compared to a 12–6 Lennard-Jones potential. In OPEPv4, authors have changed the ε_{ij} coupling constants of some interaction pairs present in α-helices based on a statistical analysis of PDB structures. To further improve OPEP, Sterpone and coworkers[89] recently derived new effective potentials for describing the ion pair interactions from all-atom PMF (potential of mean force), that

feature a single minimum for Lys-Asp and Lys-Glu pairs and two minima for Arg-Asp and Arg-Glu. The OPEP force field has been used successfully for predicting protein stabilities as well as folding and aggregation studies of small peptides[88,90] and the structure prediction of small proteins[91].

OPEP is highly transferable among different proteins, but it is limited to aqueous solvent environments. OPEP requires an integration time-step of 1.5 fs due to the detailed backbone and hydrogen bonds. This limits the amount of sampling that can be accomplished. Furthermore, a single-bead representation of the SC puts a limit on capturing the chemical specificity of SCs specificity and creates ambiguity in when reconstructing all-atom models from the CG model. Therefore, applications within multi-scale frameworks may be limited. Furthermore, due to the lack of chemical SC specificity, this model may not be suitable for investigating protein-protein interactions.

SCORPION Model

A major challenge in computational modeling of protein-protein interactions is the *de novo* prediction of protein-protein interfaces. To access the large sampling space, CG simulations are often employed. Basdevant and coworkers[14] have developed the physics-based simplified CG model SCORPION (Solvated COaRse-grained Protein interactION) which targets applications in protein-protein recognition. In SCORPION, the backbone of each amino acid is represented by a single bead while one or two beads are used to describe SCs depending on their size (Figure 2). The CG beads are located at the geometric center of the heavy atoms. This mapping scheme yields 29 CG types. An elastic network model (ENM) model was introduced in SCORPION to account for the internal protein flexibility. Because the ENM model maintains the internal protein structure, no bonded terms are necessary in SCORPION. Following classical all-atom force fields, non-polar and electrostatic interactions were expressed as sums of pairwise van der Waals and Coulombic interaction potentials and were optimized in a consistent bottom-up approach. Van der Waals interactions were extracted by fitting van der Waals SC pair interactions to PMFs extracted from all-atom MD simulations that were performed in vacuum with vanishing charges (i.e., no Coulomb interactions) in order to account for the purely non-electrostatic interaction. The resulting van der Waals interaction potential is softer than its atomistic counterpart and is composed of a very smooth repulsive part in $1/r^6$ and a Gaussian attractive part[14]. The CG point charges for a protein were obtained by reproducing the electrostatic potentials obtained with an all-atom model in vacuum. The motivation for carrying out the parameterization in vacuum was to obtain a model that can then be coupled to any implicit or explicit solvent model. While this makes sense in principle, this approach neglects solute polarization that is induced upon transfer from vacuum to solvent environments and that may require a polarizable version of SCORPION. As a step in that direction, the authors of SCORPION have recently developed a polarizable coarse-grained solvent (PCGS) model, which can capture such polarization effects in part and was validated successfully against solvation free energies of peptides. The resulting CG model was found to be about 100–130 times faster than the atomistic simulation.

The combined protein and water model was used successfully in mechanistic studies of protein-protein recognition. For instance, they have conducted seven long molecular dynamics simulations starting from several initial configurations to predict the structure of barnase/barstar complex. Notably, the simulations were carried out at a very high temperature (500 K) to accelerate the phase-space sampling. Within 1 to 500 ns, they observed that five of seven simulations adopted a conformation very similar to the native complex structure with RMSDs from the crystal structure varying between 0.5 and 2 Å. Furthermore, their simulations reveal that this protein-protein recognition process is mainly driven by the van der Waals and electrostatic interactions between two partner proteins. These results agree with the previous all-atom simulations[92].

The current model suffers from two main drawbacks: lack of bonded interactions and the use of high temperature that affects the balance between the enthalpic and entropic contributions.

PaLaCe Model

Recently, Pasi, Lavery and Ceres (PaLaCe)[93] developed a new CG model for proteins for investigating the mechanical properties of proteins. This model uses a two-tier representation of the polypeptide chains: one representation is used for bonded interactions and for backbone hydrogen bonds and involves atomic beads while a different level is used for non-bonded terms, other than main chain hydrogen bonds, and involves a pseudo atom (or CG) representation of the amino acids. One or two CG beads are used for SCs while three beads (N, Cα, and C`) are used to represent the backbone so that an explicit description of hydrogen bonds becomes possible (Figure 2). The potential energy function of PaLaCe consists of physics-based bonded and non-bonded terms with an additional term for the backbone hydrogen bonding potential, which is introduced to allow secondary structures changes. The van der Waals interactions in PaLaCe are characterized by a Lennard-Jones-like interaction but with a softer repulsive term than atomistic force fields. The Hydrogen bonding potential is composed of 12-6 Lennard-Jones and electrostatic contributions. The electrostatic component of the hydrogen bonding potential is modeled via Coulomb potential damped by a distance-dependent dielectric constant. The solvent is treated implicitly with a one-body solvation term based on circular variance, which is used to describe the occlusion of a CG atom within a protein structure. The PaLaCe potential was parameterized following a bottom-up approach by applying the Boltzmann inversion technique to conformational probability distributions derived from a large data set of protein structures followed by iterative refinement. PaLaCe was shown to maintain stable structures of folded proteins. It can also correctly model large-scale, force-induced conformational changes for the immunoglobulin-like domain of the giant protein titin. PaLaCe is reported to be faster than equivalent all-atom simulations in an explicit solvent by roughly 1000 times. The model is limited to aqueous solvent environments only.

Bereau and Deserno Model

Another transferable CG model for proteins was designed by Bereau and Deserno[94]. It uses an intermediate level of description of the peptide and treats solvent effects implicitly. Each amino acid is described by four beads: three beads for the backbone and one bead for SC.

The CG beads for backbones are located at N, Cα and C′ while the CG bead for SCs is located at Cβ (Figure 2). For non-glycine amino acids, all SC beads are assigned the same van der Waals radius. The bonded terms are generic and the corresponding geometric parameters were taken from existing peptide models[39,95,96]. The spring constants are set high enough and an approximate 5% flexibility around their reference values was given to account for thermal fluctuations. The peptide bond is modeled by a single minimum centered around the *trans* conformation while two minima centered around *cis* and *trans* conformations are employed to model the isomerization.

The SC non-bonded interactions were described by the venerable knowledge-based potential from Miyazawa and Jernigan[17] that resulted from a statistical analysis of SC contacts in a protein structure database. No explicit electrostatic interactions were included since they are implicitly accounted for in several of the other terms. The backbone bead interaction potential includes a short-range, purely repulsive Week-Chandler-Anderson (WCA) potential[97] to capture locally excluded volume interactions. Furthermore, a backbone hydrogen bonding potential as a function of the relative distance and orientation of an amide and a carbonyl group was added following Irbäck et al.[39], where a radial 12-10 Lennard-Jones potential is combined with an angular term. It has been shown previously[98] that carbonyl and amide groups at the peptide bond form dipoles that interact with each other and that the nearest-neighbor interaction is enough to favor β over α content[99]. Therefore, an extra term describing dipole-dipole interactions of neighboring residues is also included in this model. A single cosine function centered around φ, $\psi = 0$ with an identical amplitude along both coordinates was employed to model the neighboring dipole potential. These terms were optimized by reproducing dihedral angle distributions of GGG and GAG tripeptides and global folding properties of a three helix bundle. The resulting CG force field was used for studying protein folding[100] and aggregation[94]. One of the drawbacks of this model is that the stability of β-sheet structures and proteins with mixed α/β content is apparently underestimated. The model is implemented in the ESPRESSO package[101].

PRIMO Model

Very recently, we developed the intermediate resolution CG model for proteins, PRIMO[102,103]. The backbone in PRIMO is described by three CG beads (N, Cα, and combined CO) while one to five sites are used for SCs (Figure 2). A unique feature of PRIMO is that the CG beads are mapped in such a way that atomistic models can be reconstructed from the CG beads using an analytical procedure that is extremely fast and provides near-atomistic accuracy in the reconstructed models. This allows PRIMO to be tied closely to fully atomistic models in order to maximize transferability. It also positions PRIMO as an ideal CG model for multi-scale models. Like some of the other CG models, PRIMO uses virtual sites that are constructed on the fly from the CG interaction sites in order to maintain better bonding geometries.

PRIMO is largely a physics-based model with an all-atom like force field as shown in equation 3:

$$
\begin{aligned}
U^{\mathrm{PRIMO}} = & \sum_{i=1}^{N_{\mathrm{bond}}} k_i^{\mathrm{bond}}(l_i - l_{i,0})^2 + \sum_{i=1}^{N_{\mathrm{angle}}} k_i^{\mathrm{angle}}(\theta_i - \theta_{i,0})^2 + \\
& \sum_{i=1}^{N_{\mathrm{bond\text{-}spline}}} s_i^{1D}\left(l_i^{1-2}\right) + \sum_{i=1}^{N_{\mathrm{angle\text{-}spline}}} s_i^{1D}\left(l_i^{1-3}\right) + \\
& \sum_{i=1}^{N_{\mathrm{virtual\text{-}atom}}} k_i^{\mathrm{virtual\text{-}bond}}(i_i - l_{i,0})^2 + \sum_{i=1}^{N_{\mathrm{virtual\text{-}atom}}} k_i^{\mathrm{virtual\text{-}angle}}(\theta_i - \theta_{i,0})^2 + \\
& \sum_{i=1}^{N_{\mathrm{torsion}}} \sum_{j=1}^{N_i^{\mathrm{mult}}} k_{i,j}^{\mathrm{torsion}}\left(1 + cos\left(n_{i,j}\varphi_i - \varphi_{i,j,0}\right)\right) + \sum_{i=1}^{N_{\mathrm{torsionl\text{-}spline}}} s_{\mathrm{ld}}^i(l_i^{1-4}) + \\
& \sum_{i=1}^{N_{\mathrm{CMAP}}} s_i^{2D}\left(\varphi_i, \psi_i\right) + \sum_{i=1}^{N_{\mathrm{HBOND}}} s_i^{2D}\left(cos\theta, l_i^{\mathrm{N\text{-}CO}}\right) + \\
& \sum_{i=1}^{N_{atom}-1}\sum_{j=i+1}^{N_{atom}} 4\varepsilon_{\mathrm{ij}}\left[\left(\frac{\sigma_{\mathrm{ij}}}{r_{\mathrm{ij}}}\right)^{12} - \left(\frac{\sigma_{\mathrm{ij}}}{r_{\mathrm{ij}}}\right)^6\right] + \sum_{i=1}^{N_{atom}-1}\sum_{j=i+1}^{N_{atom}} \frac{q_i q_j}{4\pi\varepsilon_0 r_{\mathrm{ij}}} + \\
& \Delta G_{\mathrm{solv}}^{\mathrm{GB}} + \sum_{i=1}^{N_{\mathrm{atom}}} \gamma_i (SASA)_i
\end{aligned}
\tag{2}
$$

The bonded terms are described by harmonic potentials as well as spline potentials to model non-harmonic bonding functions between CG beads. The interaction potential also includes a distance- and angle-dependent explicit hydrogen-bonding potential term because hydrogen-bonding cannot be captured with electrostatics alone in CG models. A CMAP[104] correction term is used to provide a correct description of the backbone (φ, ψ) and SC (χ) distributions. PRIMO uses an implicit solvent potential to capture solvent effects. PRIMO was optimized by direct comparison with the latest all-atom CHARMM36 force field[105,106] against a diverse set of decoys of peptides and proteins. There is a high energetic correspondence between PRIMO and CHARMM which provides the basis for a seamless mixing of PRIMO and CHARMM in multi-scale AA/CG simulations[107]. In contrast to most other similar CG models, PRIMO does not require any bias towards known structural information and is suitable for running stable MD simulations of proteins and for folding of small peptides from their extended conformations. PRIMO allows a time step of 4–6 fs and achieves a speedup of about 10–15 times faster compared to atomistic simulations. This speedup is significantly less compared to other CG models in part due to the relatively high resolution but also because of the expensive GBMV implicit solvent model[108] that is currently used in combination with PRIMO. PRIMO essentially maintains the same level of transferability of all-atom force fields, including transferability to different environments. For example, it is possible to transfer to membrane environments by simply switching the GBMV[108] implicit solvent model for the heterogeneous dielectric generalized Born (HDGB)[109,110] implicit membrane model without further optimizations as demonstrated recently[111].

GBEMP Model

In general, CG beads are considered to be isotropic. However, in the context of biomolecular systems, CG beads are anisotropic because a single CG bead represents a group of atoms.

Motivated by this observation Shen and coworkers[112] have recently introduced a physics-based CG model, named GBEMP, based on the anisotropic Gay-Berne (GB) and point electric multipole (EMP) potentials that are used to describe the non-bonded interactions. In GBEMP, the bonded terms consist of bond, angle and dihedral potentials. Instead of using simple harmonic potentials, the bond stretching term is described by the fourth-order Taylor expansion of the Morse potential while the angle bending term adopts a sixth-order potential. A Fourier series expansion is used to model the torsional potential. The generalized Kirkwood (GK)[113] implicit solvent model was employed to describe the solvent effect in GBEMP. In an alternative to the traditional leap-frog or Verlet algorithm, Euler's rigid body integrator is used to propagate the equation of motion.

The force field was parameterized based on a combination of experimental and all-atom data. The bonded terms were parameterized by fitting to the PMFs obtained from atomistic molecular dynamics simulations of dipeptides using CHARMM27/CMAP[104] force field[114,115]. The parameters for torsional potentials were further optimized by iteratively matching to the experimental results for the distributions of backbone and SC torsions. The non-bonded terms were parameterized by reproducing the individual energy components of the atomistic force field. The model was validated by reproducing experimental backbone and SC conformations and stable MD simulations of two proteins. The resulting GBEMP CG force field is general and transferable, suitable for simulating biological macromolecules. The model is also suitable for "serial" or "parallel" multi-scale AA/CG simulations. The GBEMP model is about 50 times faster compared to atomistic force fields with point charge models.

PRIME Model

Hall and coworkers[116,117] have introduced an implicit solvent intermediate-resolution protein model, namely PRIME based on the discontinuous molecular dynamics (DMD)[118] methodology. Here, each amino acid is described by four beads: three beads for the backbone and a single bead for SC (Figure 2). The backbone beads are comprised of the united atoms NH, CαH and CO. Furthermore, SCs are distinguished into 14 groups depending on their size, charge, hydrophobicity, polarity, and potential for SC hydrogen bonding. All forces are modeled either by hard-sphere or by square well potentials with realistic diameters. For instance, a square-well potential was employed to model the interactions between hydrophobic SCs while the backbone hydrogen bonding potential was represented by a directionally dependent square-well attraction between NH and CO. The resulting force field contains 19 energy parameters. These parameters were obtained by using a perceptron-learning algorithm[119] together with a modified stochastic learning algorithm that optimizes the energy gap between 711 known native states from the PDB and decoys generated by gapless threading.

The SC representation in PRIME is too coarse to correctly capture the SC entropy and the size effect for the packing of SCs that make critical contributions to protein folding. Therefore, Ding and coworkers[120] have extended this model by incorporating one or two more additional effective SC atoms. For the β-branched amino acids (Thr, Ile, and Val), two

beads representing the two branches after Cβ were introduced while an additional bead was incorporated for bulky amino acids (Arg, Lys, and Trp).

3. Hybrid All-atom/Coarse-Grained (AA/CG) Modeling of Proteins

Generally, AA/CG approaches may be divided into three categories[121–123]: 1) Serial approaches where the information from fine-grained simulations are transferred to a CG model, and subsequently only the CG model is explored; 2) a tightly coupled approach where both levels are present simultaneously; and 3) a loosely coupled approach, where fine-grained and CG levels of resolution are alternated (see Figure 3). The first type is essentially the idea behind Boltzmann inversion[24] and force matching methods[25,121]. A typical example of the second type may be alternating simulations at the CG and all-atom levels, where the CG simulations are used to identify an ensemble of possible conformational states that are then connected via fine-grained simulations to obtain accurate conformational PMFs[124–126]. The third type would involve either resolution exchange simulations, with multiple replicas simulated at different resolutions[127–129], or hybrid models[25–27], where part of a system is represented in atomistic detail while other parts remain at the CG level[122,123,130–132]. Such hybrid approaches are most attractive for detailed mechanistic studies of specific parts of a large system where an explicit description at the atomistic level at the region of interest is vital but computational costs of simulating the entire complex in full atomistic detail are prohibitive.

In the first two approaches, a reverse mapping[133,134] scheme is employed once at the end (type 1) or infrequently (type 3) to reconstruct the high-resolution model from the corresponding low-resolution model. In the case of tightly coupled AA/CG simulations, the conversion between AA and CG regions is more critical as it has to be done frequently and has to maintain detailed balance when going back and forth between different resolutions. In the case of hybrid AA/CG models, the treatment at the boundary between the different regions becomes a key issue. The boundary may be treated as rigid (fixed-resolution AA/CG method)[135–138] analogous to the conventional hybrid quantum mechanics/molecular mechanics (QM/MM)[139,140] approach or it may be open (adaptive AA/CG method)[132,141]. In the fixed-resolution AA/CG scheme, the coupling is continuous in time, and the atomistic and CG components of the system directly interact. A limiting case of the adaptive AA/CG scheme are the resolution-exchange methods already mentioned above.

In this section, we will focus primarily on tightly-coupled hybrid AA/CG models as they promise to be the most effective multi-scale scheme in the foreseeable future. For an extended review on the multi-scale modeling of proteins, readers are referred to recent reviews[123,142–145].

PACE Model

Multi-scale AA/CG methods are developed to reduce the computational cost while capturing the relevant interactions at the atomic level. Since the solvent degrees of freedom are much larger compared to the solute (protein) degrees of freedom, a significant speedup can be achieved already just by representing the solvent at the CG level while the solute is described at the atomistic level. Motivated by this idea, Schulten and coworkers have

proposed PACE (Protein in Atomistic details coupled with Coarse-grained Environment)[146]. More specifically PACE combines a united-atom protein model with solvent, water or lipids, represented with the MARTINI model. The potential energy function of PACE is given by the following equation.

$$U = E_{\text{bond}} + E_{\text{angle}} + E_{\text{dihedral}} + E_{\text{improper}} + E_{\varphi,\psi,\chi_1} + E_{\text{CGW-CGW}} + E_{\text{CGW-UA}} + E_{\text{vdW}} + E_{\text{polar}} \quad (3)$$

The first four terms are generic bonded interaction potentials while the fifth term accounts for backbone and SC rotamers. The remaining terms capture non-bonded interactions that consist of standard Lennard-Jones 12-6 potentials for $E_{\text{CGW-CGW}}$ interactions between CG waters, for $E_{\text{CGW-UA}}$ interactions between CG water and protein sites, and for E_{vdW} interactions between non-polar protein sites. The interactions between charged groups are modeled with a Coulomb potential using partial charges from the all-atom force field. The cross-resolution parameters were optimized by following the parameterization scheme of MARTINI, i.e., by reproducing experimental thermodynamic properties. Moreover, in the context of the hybrid force field, the interactions of atomic sites in proteins were reparameterized by using different reference data, such as PMF of polar interactions from atomistic simulations and statistical backbone potentials from a Protein Data Bank (PDB) coil library[146]. The resulting hybrid model was able to reproduce the statistical backbone and SC potentials of all amino acids accurately. The force field was further validated with *ab initio* folding simulations of a series of peptides and proteins. For instance, PACE[146] has been employed to fold a 73-residue protein (α3D)[147], which is a *de novo* designed three-helix-bundle protein. Starting from the denatured state, Han and Schulten[146] folded this protein to an average Cα RMSD of 3.4 Å in three extended (10–30 μs) simulations. This result is very close to the result (3.1 Å) obtained by Shaw and coworkers in fully atomistic simulations[148].

While replacing only the solvent with a CG representation limits the computational gains that can be realized, this approach competes with already established implicit solvent models. Furthermore, the initial tests indicate that the stability of native proteins may be underestimated in PACE, suggesting that the balance between the atomistic protein model and CG solvent may be challenging. Nevertheless, recent studies have extended PACE further towards protein-membrane systems[149].

ATTRACT Model

Zacharias et al. recently developed a multi-scale AA/CG model[150] by combing a modified GROMOS united atom force field with the ATTRACT[151] CG model for studying protein-protein interactions. In this model, bonded interactions are generic and are described by atomistic force field while non-bonded interactions are calculated via CG terms. Backbone atoms were retained in full atomic detail to allow accurate representations of hydrogen-bonded secondary structures and an explicit term for hydrogen bonds of main chain was incorporated to maintain a better secondary structures. In the multi-scale scheme, all SC heavy atoms were treated as virtual atoms and non-bonded SC-SC as well as main chain-SC interactions were calculated via the ATTRACT CG model instead of standard atomistic

Lennard-Jones and Coulomb terms. In ATTRACT, the backbone is represented by two pseudoatoms (N and O) and SCs are represented by one or two pseudoatoms depending on their size. Effective interactions between pseudo atoms are described by soft distance-dependent Lennard-Jones-type potentials. The attractive and repulsive parameters for each pseudo atom pair were iteratively optimized based on predictions of protein–protein complexes[151]. Hence, the parameterization of this model represents the effective interactions between pseudo centers at protein-protein interfaces. The parameters implicitly account for solvation effects, and no explicit solvent molecules were included during simulations. Promising results were obtained for protein-protein interactions with this hybrid methodology. However, the current scheme is of limited use for studies of internal protein dynamics. Although ATTRACT can be used to investigate the dynamics of protein-protein recognition process, computational efficiency is an issue.

Promising Hybrid Models

Rzepiela and co-workers[152] have proposed a mixed model between all-atom and MARTINI CG force fields where all-atom-CG interactions are replaced by CG interactions through dummy sites. This avoids the need for reparameterization of cross-resolution terms from the CG perspective but it does not fully address the environment seen by the atomistic part. Following this protocol, Sokkar et al.[153] found that the MARTINI force field couples too strongly with the atomistic model leading to complete unfolding of a test protein within a very short time.

In a similar effort, the Carloni group[154] has also developed a method in which the biologically relevant region of the protein is represented at the atomic level, while the rest of the protein is treated at the CG level, by only considering Cα centroids. The atomistic region is treated using a standard molecular force field, the CG region using a simplified Go potential and the interface region is also treated with an AA force field. This AA/CG scheme shows promising results for the HIV-1 protease (HIV-1 PR) and human β-secretase (BACE). As solvent-protein interactions are modeled in terms of viscosity and environment random forces in the framework of stochastic dynamics, this current model is only suitable for simulating globular proteins with buried active sites. Further improvements are necessary for modeling the solvent flow in the AA part.

The Pantano group has recently introduced a CG model for DNA[155] and later combined it with the atomistic AMBER force field[156] for multi-scale AA/CG simulations of DNA[157]. Villa and coworkers developed a multi-scale method for simulating protein-DNA complexes[158] and conducted a multi-scale simulation to investigate a complex between the lac repressor protein (LacI) and a 107 bp DNA segment[159]. The LacI complex with the operator DNA segment is modeled at the level of atomic details while a continuum model is employed for the 75 bp DNA loop. Using this multi-scale strategy, they have investigated the change in structural dynamics of the repressor due to the strained DNA loop.

Finally, PRIMO[103] is another promising CG model that can be used to carry out multi-scale AA/CG simulations of proteins. Recently, Predeus and coworkers[107] have employed PRIMO[103] and CHARMM[114] force fields simultaneously for conformational sampling of trp-cage and melittin peptides in the presence of crowder proteins. A three-component multi-

scale scheme was adopted in their simulations. The peptides of interest were modeled with the atomistic CHARMM force field[114] while the PRIMO force field was employed to model the crowder proteins and the solvent was treated implicitly via the GBMV[108] scheme. Their simulations yielded significant populations of unfolded structures of trp-cage and melittin in the presence of cellular crowders. It should be noted here that there were no direct bonded interactions between AA and CG regions in this multi-scale AA/CG hybrid model. Both regions were interacting with each other via non-bonded interactions only. However, PRIMO can also be coupled with the atomistic force field to describe the direct bonded interactions between AA and CG atoms with minimum effort. The coupling can be achieved via two ways as shown in Figures 4 A and B. A crude way of combing PRIMO with CHARMM is to let the CG sites directly interact with the atomistic sites. One such example is shown in Figure 4C. Here, 1–14 residues of trp-cage are modeled with the CHARMM force filed while residues from 15–20 are described by the PRIMO force field. The cross-resolution parameters can be taken from the existing PRIMO or CHARMM model as shown in Figure 4D. An advantage of such a mixing scheme is that it avoids an extensive parameterization of the interfacial parameters.

A sophisticated way of achieving multi-scale coupling is shown in Figure 4B. Here, the AA part is converted to CG on the fly and let it interacts with the corresponding CG site. Furthermore, on the CG side, virtual sites are constructed on the fly and let it interacts with the corresponding atomistic bead from the AA part. This is a very robust approach and can be easily implemented as the reverse mapping in PRIMO allows a near atomistic accuracy of 0.1 Å.

4. Outlook

During the last years we have seen the development of numerous CG models that are developed to study various biophysical processes. In this chapter, we have focused on those CG models that maximize transferability and are thus more universally applicable. We also discuss extensively multi-scale hybrid AA/CG methods that rely on transferable CG models and are starting to become useful for practical applications. While we believe, that the use of CG models and hybrid AA/CG schemes will continue to grow in the future many limitations remain as described extensively, e.g. by Saunders and Voth[34]. Still, even most 'transferable' CG models currently available are biased towards known data and therefore suffer from limitations when applied to a full range of current problems (protein dynamics, protein-protein interactions, folding etc.), when considering different environments, when combined with different biomolecules, or under varying thermodynamic conditions. One question how to further address this challenge is how to balance bottom-up and top-down approaches in developing effective CG models as both have strengths and weaknesses. A mixed approach was adopted in the development of MARTINI force field with overall good success. However, there are also fundamental questions of what degree transferability is theoretically possible at a chosen level of resolution[34,160].

The coupling of AA and CG models in hybrid multi-scale schemes is still in its infancies. While first success stories have been reported, a robust strategy for tightly coupling AA and CG levels across an interface or between replicas at different resolutions is lacking. One of

the outstanding issues is the preservation of detailed balance at both the AA and CG levels. Other questions are at what levels CG and AA models can be effectively combined since CG models that are too coarse are most certainly going to lead to problematic behavior when coupled directly to detailed atomistic potential.

Another topic that has remained largely unaddressed is how the dynamics resulting from CG and hybrid AA/CG models reproduces the correct kinetic behavior. For thermodynamic questions this can be neglected but as CG models are increasingly applied to cellular scales[107,161,162] questions related to diffusive properties become more relevant[163,164]. While it is expected that CG models generally exhibit accelerated kinetics, it is much less clear whether different types of motion are affected similarly. Even more problematic is how dynamic properties may be affected when mixing AA and CG models. Therefore, a future challenge will be to develop a CG model that is dynamically accurate as well as computationally efficient[34].

5. Conclusions

In this chapter, we have reviewed recent advances in developing transferable, more universally applicable CG models of proteins. While there is clear progress in recent years, many challenges remain. Another promising area is the development of multi-scale AA/CG hybrid methods that may benefit most from more transferable CG models. It is likely that with further increased accuracy and efficiency, CG and AA/CG schemes will play an even greater role in the future in addressing questions related to biomolecular dynamics and interactions on ever increasing spatiotemporal scales all the way to cellular scales.

Acknowledgments

Bibliography

1. Karplus M, McCammon JA. Molecular dynamics simulations of biomolecules. Nat Struct Biol. 2002; 9:646–652. [PubMed: 12198485]

2. Freddolino PL, Liu F, Gruebele M, Schulten K. Ten-microsecond molecular dynamics simulation of a fast-folding WW domain. Biophys J. 2008; 94:L75–L77. [PubMed: 18339748]

3. 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]

4. Harada R, Sugita Y, Feig M. Protein Crowding Affects Hydration Structure and Dynamics. J Am Chem Soc. 2012; 134:4842–4849. [PubMed: 22352398]

5. Harada R, Tochio N, Kigawa T, Sugita Y, Feig M. Reduced Native State Stability in Crowded Cellular Environment Due to Protein-Protein Interactions. J Am Chem Soc. 2013; 135:3696–3701. [PubMed: 23402619]

6. Feig M, Sugita Y. Protein-Protein Interactions in Crowded Cellular Environments. Biophys J. 2012; 102:473a–473a.

7. 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]

8. Law SM, Feig M. Base-Flipping Mechanism in Postmismatch Recognition by MutS. Biophys J. 2011; 101:2223–2231. [PubMed: 22067162]

9. Sharma MPAV, Kovacs N, Feig M. MutS Differential mismatch recognition specificities of eukaryotic MutS homologα, MutSa and MutSβ. Biophys J. 2014 In press.

10. Majek P, Elber R. A coarse-grained potential for fold recognition and molecular dynamics simulations of proteins. Proteins. 2009; 76:822–836. [PubMed: 19291741]

11. Tozzini V. Coarse-grained models for proteins. Curr Opin Struc Biol. 2005; 15:144–150.

12. Maisuradze GG, Senet P, Czaplewski C, Liwo A, Scheraga HA. Investigation of Protein Folding by Coarse-Grained Molecular Dynamics with the UNRES Force Field. J Phys Chem A. 2010; 114:4471–4485. [PubMed: 20166738]

13. Thorpe IF, Zhou J, Voth GA. Peptide folding using multiscale coarse-grained models. J Phys Chem B. 2008; 112:13079–13090. [PubMed: 18808094]

14. Basdevant N, Borgis D, Ha-Duong T. Modeling Protein-Protein Recognition in Solution Using the Coarse-Grained Force Field SCORPION. J Chem Theory Comput. 2013; 9:803–813. [PubMed: 26589072]

15. Kolinski A. Protein modeling and structure prediction with a reduced representation. Acta Biochim Pol. 2004; 51:349–371. [PubMed: 15218533]

16. Maupetit J, Tuffery P, Derreumaux P. A coarse-grained protein force field for folding and structure prediction. Proteins. 2007; 69:394–408. [PubMed: 17600832]

17. Miyazawa S, Jernigan RL. Estimation of Effective Interresidue Contact Energies from Protein Crystal-Structures - Quasi-Chemical Approximation. Macromolecules. 1985; 18:534–552.

18. Krishna V, Ayton GS, Voth GA. Role of Protein Interactions in Defining HIV-1 Viral Capsid Shape and Stability: A Coarse-Grained Analysis. Biophys J. 2010; 98:18–26. [PubMed: 20085716]

19. Arkhipov A, Freddolino PL, Imada K, Namba K, Schulten K. Coarse-grained molecular dynamics simulations of a rotating bacterial flagellum. Biophys J. 2006; 91:4589–4597. [PubMed: 16997871]

20. Ayton GS, Voth GA. Systematic multiscale simulation of membrane protein systems. Curr Opin Struc Biol. 2009; 19:138–144.

21. Ayton GS, Voth GA. Multiscale Computer Simulation of the Immature HIV-1 Virion. Biophys J. 2010; 99:2757–2765. [PubMed: 21044572]

22. Sinitskiy AV, Saunders MG, Voth GA. Optimal Number of Coarse-Grained Sites in Different Components of Large Biomolecular Complexes. J Phys Chem B. 2012; 116:8363–8374. [PubMed: 22276676]

23. Lyubartsev AP, Laaksonen A. Calculation of Effective Interaction Potentials from Radial-Distribution Functions - a Reverse Monte-Carlo Approach. Phys Rev E. 1995; 52:3730–3737.

24. Reith D, Putz M, Muller-Plathe F. Deriving effective mesoscale potentials from atomistic simulations. J Comput Chem. 2003; 24:1624–1636. [PubMed: 12926006]

25. Izvekov S, Voth GA. A multiscale coarse-graining method for biomolecular systems. J Phys Chem B. 2005; 109:2469–2473. [PubMed: 16851243]

26. Noid WG, Chu JW, Ayton GS, Krishna V, Izvekov S, Voth GA, Das A, Andersen HC. The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models. J Chem Phys. 2008; 128

27. Noid WG, Liu P, Wang Y, Chu JW, Ayton GS, Izvekov S, Andersen HC, Voth GA. The multiscale coarse-graining method. II. Numerical implementation for coarse-grained molecular models. J Chem Phys. 2008; 128

28. Chu JW, Izvekov S, Voth GA. The multiscale challenge for biomolecular systems: coarse-grained modeling. Mol Simulat. 2006; 32:211–218.

29. Chaimovich A, Shell MS. Coarse-graining errors and numerical optimization using a relative entropy framework. J Chem Phys. 2011; 134

30. Brini E, Herbers CR, Deichmann G, van der Vegt NFA. Thermodynamic transferability of coarse-grained potentials for polymer-additive systems. Phys Chem Chem Phys. 2012; 14:11896–11903. [PubMed: 22722404]

31. Brini E, van der Vegt NFA. Chemically transferable coarse-grained potentials from conditional reversible work calculations. J Chem Phys. 2012; 137

32. Henderso.Rl. Uniqueness Theorem for Fluid Pair Correlation- Functions. Phys Lett A. 1974; A 49:197–198.

33. Ingolfsson HILCA, Uusitalo JJ, de Jong DH, Gopal S, Periole X, Marrink SJ. The power of coarse-graining in biomolecular simulations. WIREs Comput Mol Sci. 2014; 4:225–248.

34. Saunders MG, Voth GA. Coarse-Graining Methods for Computational Biology. Annu Rev Biophys. 2013; 42:73–93. [PubMed: 23451897]

35. Levitt M, Warshel A. Computer-Simulation of Protein Folding. Nature. 1975; 253:694–698. [PubMed: 1167625]

36. Buchete NV, Straub JE, Thirumalai D. Continuous anisotropic representation of coarse-grained potentials for proteins by spherical harmonics synthesis. J Mol Graph Model. 2004; 22:441–450. [PubMed: 15099839]

37. Brown S, Fawzi NJ, Head-Gordon T. Coarse-grained sequences for protein folding and design. P Natl Acad Sci USA. 2003; 100:10712–10717.

38. Tozzini V, McCammon JA. The dynamics of the flap opening in HIV- protease: A coarse grained approach. Protein Sci. 2004; 13:194–194.

39. Irback A, Sjunnesson F, Wallin S. Three-helix-bundle protein in a Ramachandran model. P Natl Acad Sci USA. 2000; 97:13614–13618.

40. Fernandez A, Colubri A. Pathway heterogeneity in protein folding. Proteins-Structure Function and Genetics. 2002; 48:293–310.

41. Fujitsuka Y, Takada S, Luthey-Schulten ZA, Wolynes PG. Optimizing physical energy functions for protein folding. Proteins-Structure Function and Genetics. 2004; 54:88–103.

42. Ghavami A, van der Giessen E, Onck PR. Coarse-Grained Potentials for Local Interactions in Unfolded Proteins. J Chem Theory Comput. 2013; 9:432–440. [PubMed: 26589045]

43. Bahar I, Jernigan RL. Inter-residue potentials in globular proteins and the dominance of highly specific hydrophilic interactions at close separation. J Mol Biol. 1997; 266:195–214. [PubMed: 9054980]

44. Honeycutt JD, Thirumalai D. Metastability of the Folded States of Globular-Proteins. P Natl Acad Sci USA. 1990; 87:3526–3529.

45. Korkuta A, Hendrickson WA. A force field for virtual atom molecular mechanics of proteins. P Natl Acad Sci USA. 2009; 106:15667–15672.

46. Izvekov S, Chung PW, Rice BM. The multiscale coarse-graining method: Assessing its accuracy and introducing density dependent coarse-grain potentials. J Chem Phys. 2010; 133

47. Izvekov S, Voth GA. Multiscale coarse graining of liquid-state systems. J Chem Phys. 2005; 123

48. Hills RD Jr, Lu L, Voth GA. Multiscale coarse-graining of the protein energy landscape. PLoS computational biology. 2010; 6:e1000827. [PubMed: 20585614]

49. Head-Gordon T, Brown S. Minimalist models for protein folding and design. Curr Opin Struc Biol. 2003; 13:160–167.

50. Takada S. Coarse-grained molecular simulations of large biomolecules. Curr Opin Struc Biol. 2012; 22:130–137.

51. Tozzini V. Minimalist models for proteins: a comparative analysis. Q Rev Biophys. 2010; 43:333–371. [PubMed: 20707940]

52. Mukherjee A, Bagchi B. Contact pair dynamics during folding of two small proteins: Chicken villin head piece and the Alzheimer protein beta-amyloid. J Chem Phys. 2004; 120:1602–1612. [PubMed: 15268287]

53. Liwo A, Oldziej S, Kazmierkiewicz R, Groth M, Czaplewski C. Design of a knowledge-based force field for off-lattice simulations of protein structure. Acta Biochim Pol. 1997; 44:527–547. [PubMed: 9511963]

54. Liwo A, Oldziej S, Pincus MR, Wawak RJ, Rackovsky S, Scheraga HA. A united-residue force field for off-lattice protein-structure simulations .1. Functional forms and parameters of long-range side-chain interaction potentials from protein crystal data. J Comput Chem. 1997; 18:849–873.

55. Liwo A, Pincus MR, Wawak RJ, Rackovsky S, Oldziej S, Scheraga HA. A united-residue force field for off-lattice protein-structure simulations .2. Parameterization of short-range interactions

and determination of weights of energy terms by Z-score optimization. J Comput Chem. 1997; 18:874–887.

56. Khalili M, Liwo A, Rakowski F, Grochowski P, Scheraga HA. Molecular dynamics with the united-residue model of polypeptide chains. I. Lagrange equations of motion and tests of numerical stability in the microcanonical mode. J Phys Chem B. 2005; 109:13785–13797. [PubMed: 16852727]

57. Maisuradze GG, Liwo A, Scheraga HA. Relation between Free Energy Landscapes of Proteins and Dynamics. J Chem Theory Comput. 2010; 6:583–595. [PubMed: 23620713]

58. Liwo A, Arlukowicz P, Czaplewski C, Oldziej S, Pillardy J, Scheraga HA. A method for optimizing potential-energy functions by a hierarchical design of the potential-energy landscape: Application to the UNRES force field. P Natl Acad Sci USA. 2002; 99:1937–1942.

59. Liwo A, Khalili M, Scheraga HA. Ab initio simulations of protein folding pathways by molecular dynamics with the united-residue (UNRES) model of polypeptide chains. Febs J. 2005; 272:359–360.

60. Lee J, Liwo A, Scheraga HA. Energy-based de novo protein folding by conformational space annealing and an off-lattice united-residue force field: Application to the 10–55 fragment of staphylococcal protein A and to apo calbindin D9K. P Natl Acad Sci USA. 1999; 96:2025–2030.

61. Oldziej S, Czaplewski C, Liwo A, Chinchio M, Nanias M, Vila JA, Khalili M, Arnautova YA, Jagielska A, Makowski M, Schafroth HD, Kazmierkiewicz R, Ripoll DR, Pillardy J, Saunders JA, Kang YK, Gibson KD, Scheraga HA. Physics-based protein-structure prediction using a hierarchical protocol based on the UNRES force field: Assessment in two blind tests. P Natl Acad Sci USA. 2005; 102:7547–7552.

62. Nanias M, Chinchio M, Oldziej S, Czaplewski C, Scheraga HA. Protein structure prediction with the UNRES force-field using Replica-Exchange Monte Carlo-with-Minimization; Comparison with MCM, CSA, and CFMC. J Comput Chem. 2005; 26:1472–1486. [PubMed: 16088925]

63. Solernou A, Fernandez-Recio J. pyDockCG: New Coarse-Grained Potential for Protein Protein Docking. J Phys Chem B. 2011; 115:6032–6039. [PubMed: 21506617]

64. Rojas AV, Liwo A, Scheraga HA. Molecular dynamics with the united-residue force field: Ab initio folding simulations of multichain proteins. J Phys Chem B. 2007; 111:293–309. [PubMed: 17201452]

65. 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]

66. Marrink SJ, de Vries AH, Mark AE. Coarse grained model for semiquantitative lipid simulations. J Phys Chem B. 2004; 108:750–760.

67. Marrink SJ, Risselada HJ, Yefimov S, Tieleman DP, de Vries AH. The MARTINI force field: Coarse grained model for biomolecular simulations. J Phys Chem B. 2007; 111:7812–7824. [PubMed: 17569554]

68. Marrink SJ, de Vries AH, Harroun TA, Katsaras J, Wassall SR. Cholesterol shows preference for the interior of polyunsaturated lipid membranes. J Am Chem Soc. 2008; 130:10–11. [PubMed: 18076174]

69. Marrink SJ, de Vries AH, Tieleman DP. Lipids on the move: simulations of membrane pores, domains, stalks and curves. Biochimica et biophysica acta. 2009; 1788:149–168. [PubMed: 19013128]

70. Periole X, Huber T, Marrink SJ, Sakmar TP. G protein-coupled receptors self-assemble in dynamics simulations of model bilayers. J Am Chem Soc. 2007; 129:10126–10132. [PubMed: 17658882]

71. Risselada HJ, Marrink SJ. The molecular face of lipid rafts in model membranes. Proceedings of the National Academy of Sciences of the United States of America. 2008; 105:17367–17372. [PubMed: 18987307]

72. Sengupta D, Leontiadou H, Mark AE, Marrink SJ. Molecular dynamics simulations of toroidal pores formed by antimicrobial peptides. Biophys J. 2007:545A–545A.

73. Treptow W, Marrink SJ, Tarek M. Gating motions in voltage-gated potassium channels revealed by coarse-grained molecular dynamics simulations. J Phys Chem B. 2008; 112:3277–3282. [PubMed: 18293960]

74. Smirnova YG, Marrink SJ, Lipowsky R, Knecht V. Solvent-exposed tails as prestalk transition states for membrane fusion at low hydration. J Am Chem Soc. 2010; 132:6710–6718. [PubMed: 20411937]

75. Periole X, Knepp AM, Sakmar TP, Marrink SJ, Huber T. Structural Determinants of the Supramolecular Organization of G Protein-Coupled Receptors in Bilayers. J Am Chem Soc. 2012; 134:10959–10965. [PubMed: 22679925]

76. Marrink SJ, Tieleman DP. Perspective on the Martini model. Chem Soc Rev. 2013; 42:6801–6822. [PubMed: 23708257]

77. Yesylevskyy SO, Schafer LV, Sengupta D, Marrink SJ. Polarizable Water Model for the Coarse-Grained MARTINI Force Field. PLoS computational biology. 2010; 6

78. Wu Z, Cui QA, Yethiraj A. A New Coarse-Grained Model for Water: The Importance of Electrostatic Interactions. J Phys Chem B. 2010; 114:10524–10529. [PubMed: 20701383]

79. de Jong DH, Singh G, Bennett WFD, Arnarez C, Wassenaar TA, Schafer LV, Periole X, Tieleman DP, Marrink SJ. Improved Parameters for the Martini Coarse-Grained Protein Force Field. J Chem Theory Comput. 2013; 9:687–697. [PubMed: 26589065]

80. Singh G, Tieleman DP. Using the Wimley-White Hydrophobicity Scale as a Direct Quantitative Test of Force Fields: The MARTINI Coarse-Grained Model. J Chem Theory Comput. 2011; 7:2316–2324. [PubMed: 26606499]

81. de Jong DH, Periole X, Marrink SJ. Dimerization of Amino Acid Side Chains: Lessons from the Comparison of Different Force Fields. J Chem Theory Comput. 2012; 8:1003–1014. [PubMed: 26593362]

82. Gu JF, Bai F, Li HL, Wang XC. A Generic Force Field for Protein Coarse-Grained Molecular Dynamics Simulation. Int J Mol Sci. 2012; 13:14451–14469. [PubMed: 23203075]

83. Spiga E, Alemani D, Degiacomi MT, Cascella M, Dal Peraro M. Electrostatic-Consistent Coarse-Grained Potentials for Molecular Simulations of Proteins. J Chem Theory Comput. 2013; 9:3515–3526. [PubMed: 26584108]

84. Derreumaux P. From polypeptide sequences to structures using Monte Carlo simulations and an optimized potential. J Chem Phys. 1999; 111:2301–2310.

85. Barducci A, Bonomi M, Derreumaux P. Assessing the Quality of the OPEP Coarse-Grained Force Field. J Chem Theory Comput. 2011; 7:1928–1934. [PubMed: 26596453]

86. Sterpone FMS, Tuffery P, Pasquali S, Mousseau N, Cragnolini T, Chebaro Y, St-Pierre J-F, Kalimeri M, Barducci A, Laurin Y, Tek A, Baaden M, Nguyen PH, Derreumaux P. The OPEP protein model: from single molecules, amyloid formation, crowding and hydrodynamics to DNA/RNA systems. Chemical Society Review. 2014

87. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. A 2nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic- Molecules. J Am Chem Soc. 1995; 117:5179–5197.

88. Chebaro Y, Pasquali S, Derreumaux P. The Coarse-Grained OPEP Force Field for Non-Amyloid and Amyloid Proteins. J Phys Chem B. 2012; 116:8741–8752. [PubMed: 22742737]

89. Sterpone F, Nguyen PH, Kalimeri M, Derreumaux P. Importance of the Ion-Pair Interactions in the OPEP Coarse-Grained Force Field: Parametrization and Validation. J Chem Theory Comput. 2013; 9:4574–4584. [PubMed: 25419192]

90. Chebaro Y, Mousseau N, Derreumaux P. Structures and Thermodynamics of Alzheimer's Amyloid-beta A beta(16-35) Monomer and Dimer by Replica Exchange Molecular Dynamics Simulations: Implication for Full-Length A beta Fibrillation. J Phys Chem B. 2009; 113:7668–7675. [PubMed: 19415895]

91. Maupetit J, Derreumaux P, Tuffery P. PEP-FOLD: an online resource for de novo peptide structure prediction. Nucleic Acids Res. 2009; 37:W498–W503. [PubMed: 19433514]

92. Hoefling M, Gottschalk KE. Barnase-Barstar: From first encounter to final complex. J Struct Biol. 2010; 171:52–63. [PubMed: 20211732]

93. Pasi M, Lavery R, Ceres N. PaLaCe: A Coarse-Grain Protein Model for Studying Mechanical Properties. J Chem Theory Comput. 2013; 9:785–793. [PubMed: 26589071]

94. Bereau T, Deserno M. Generic coarse-grained model for protein folding and aggregation. J Chem Phys. 2009; 130

95. Ding F, Borreguero JM, Buldyrey SV, Stanley HE, Dokholyan NV. Mechanism for the alpha-helix to beta-hairpin transition. Proteins-Structure Function and Genetics. 2003; 53:220–228.

96. Takada S, Luthey-Schulten Z, Wolynes PG. Folding dynamics with nonadditive forces: A simulation study of a designed helical protein and a random heteropolymer. J Chem Phys. 1999; 110:11616–11629.

97. Weeks JDCD, Andersen HC. Role of repulsive forces in determining the equilibrium structure of simple liquids. J Chem Phys. 1971; 54:5237–5247.

98. Chen NY, Su ZY, Mou CY. Effective potentials for folding proteins. Phys Rev Lett. 2006; 96

99. Mu Y, Gao YQ. Effects of hydrophobic and dipole-dipole interactions on the conformational transitions of a model polypeptide. J Chem Phys. 2007; 127

100. Bereau T, Bachmann M, Deserno M. Interplay between Secondary and Tertiary Structure Formation in Protein Folding Cooperativity. J Am Chem Soc. 2010; 132:13129–13131. [PubMed: 20822175]

101. Limbach HJAA, Mann BA, Holm C. ESPRESSO-an extensible simulation package for research on soft matter systems. Comput Phys Commun. 2006; 174:704–727.

102. 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]

103. Kar P, Gopal SM, Cheng YM, Predeus A, Feig M. PRIMO: A Transferable Coarse-Grained Force Field for Proteins. J Chem Theory Comput. 2013; 9:3769–3788. [PubMed: 23997693]

104. MacKerell AD, Feig M, Brooks CL. Improved treatment of the protein backbone in empirical force fields. J Am Chem Soc. 2004; 126:698–699. [PubMed: 14733527]

105. 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 alpha-Helix and beta-Hairpin Formation. Biophys J. 2012; 103:1045–1051. [PubMed: 23009854]

106. Best RB, Zhu X, Shim J, Lopes PEM, Mittal J, Feig M, MacKerell AD. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone phi, psi and Side-Chain chi(1) and chi(2) Dihedral Angles. J Chem Theory Comput. 2012; 8:3257–3273. [PubMed: 23341755]

107. 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]

108. Lee MS, Feig M, Salsbury FR, Brooks CL. New analytic approximation to the standard molecular volume definition and its application to generalized born calculations (vol 24, pg 1348, 2003). J Comput Chem. 2003; 24:1821–1821.

109. Tanizaki S, Feig M. A generalized Born formalism for heterogeneous dielectric environments: Application to the implicit modeling of biological membranes. J Chem Phys. 2005; 122

110. 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]

111. Kar PGSM, Cheng Y-M, Panahi A, Feig M. Transferring the PRIMO coarse-grained force field to the membrane environment: Simulations of membrane proteins and helix-helix association. J Chem Theory Comput. 2014 Accepted.

112. Shen HJ, Li Y, Ren PY, Zhang DL, Li GH. Anisotropic Coarse-Grained Model for Proteins Based On Gay-Berne and Electric Multipole Potentials. J Chem Theory Comput. 2014; 10:731–750. [PubMed: 24659927]

113. Schnieders MJ, Ponder JW. Polarizable atomic multipole solutes in a generalized Kirkwood continuum. J Chem Theory Comput. 2007; 3:2083–2097. [PubMed: 26636202]

114. 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]

115. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B. 1998; 102:3586–3616. [PubMed: 24889800]

116. Cheon M, Chang I, Hall CK. Extending the PRIME model for protein aggregation to all 20 amino acids. Proteins. 2010; 78:2950–2960. [PubMed: 20740494]

117. Smith AV, Hall CK. Bridging the gap between homopolymer and protein models: A discontinuous molecular dynamics study. J Chem Phys. 2000; 113:9331–9342.

118. Dokholyan NV, Buldyrev SV, Stanley HE, Shakhnovich EI. Discrete molecular dynamics studies of the folding of a protein-like model. Fold Des. 1998; 3:577–587. [PubMed: 9889167]

119. Krauth W, Mezard M. Learning Algorithms with Optimal Stability in Neural Networks. J Phys a-Math Gen. 1987; 20:L745–L752.

120. Ding F, Buldyrev SV, Dokholyan NV. Folding Trp-cage to NMR resolution native structure using a coarse-grained protein model. Biophys J. 2005; 88:147–155. [PubMed: 15533926]

121. Ayton GS, Noid WG, Voth GA. Multiscale modeling of biomolecular systems: in serial and in parallel. Curr Opin Struc Biol. 2007; 17:192–198.

122. Kamerlin SCL, Vicatos S, Dryga A, Warshel A. Coarse-Grained (Multiscale) Simulations in Studies of Biophysical and Chemical Systems. Annu Rev Phys Chem. 2011; 62:41–64. [PubMed: 21034218]

123. Kamerlin SCL, Warshel A. Multiscale modeling of biological functions. Phys Chem Chem Phys. 2011; 13:10401–10411. [PubMed: 21526232]

124. Fujisaki H, Shiga M, Kidera A. A Multi Scale Approach for Path Sampling: Applications to Peptides. Biophys J. 2012; 102:22a–22a.

125. Fujisaki H, Shiga M, Moritsugu K, Kidera A. Multiscale enhanced path sampling based on the Onsager-Machlup action: Application to a model polymer. J Chem Phys. 2013; 139

126. 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]

127. Lyman E, Ytreberg FM, Zuckerman DM. Resolution exchange simulation. Phys Rev Lett. 2006; 96

128. Lyman E, Zuckerman DM. Resolution exchange simulation with incremental coarsening. J Chem Theory Comput. 2006; 2:656–666. [PubMed: 26626672]

129. Liu P, Voth GA. Smart resolution replica exchange: An efficient algorithm for exploring complex energy landscapes. J Chem Phys. 2007; 126

130. Murtola T, Bunker A, Vattulainen I, Deserno M, Karttunen M. Multiscale modeling of emergent materials: biological and soft matter. Phys Chem Chem Phys. 2009; 11:1869–1892. [PubMed: 19279999]

131. Peter C, Kremer K. Multiscale simulation of soft matter systems - from the atomistic to the coarse-grained level and back. Soft Matter. 2009; 5:43574366.

132. Praprotnik M, Delle Site L, Kremer K. Multiscale simulation of soft matter: From scale bridging to adaptive resolution. Annual Review of Physical Chemistry. 2008; 59:545–571.

133. Cheng YM, Gopal SM, Law SM, Feig M. Molecular Dynamics Trajectory Compression with a Coarse-Grained Model. Ieee Acm T Comput Bi. 2012; 9:476–486.

134. Wassenaar TA, Pluhackova K, Bockmann RA, Marrink SJ, Tieleman DP. Going Backward: A Flexible Geometric Approach to Reverse Transformation from Coarse Grained to Atomistic Models. J Chem Theory Comput. 2014; 10:676–690. [PubMed: 26580045]

135. 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]

136. Neri M, Anselmi C, Carnevale V, Vargiu AV, Carloni P. Molecular dynamics simulations of outer-membrane protease T from E-coli based on a hybrid coarse-grained/atomistic potential. J Phys-Condens Mat. 2006; 18:S347–S355.

137. Michel J, Orsi M, Essex JW. Prediction of partition coefficients by multiscale hybrid atomic-level/coarse-grain simulations. J Phys Chem B. 2008; 112:657–660. [PubMed: 18163606]

138. Orsi M, Sanderson WE, Essex JW. Permeability of Small Molecules through a Lipid Bilayer: A Multiscale Simulation Study. J Phys Chem B. 2009; 113:12019–12029. [PubMed: 19663489]

139. Warshel A. Computer simulations of enzyme catalysis: Methods, progress, and insights. Annu Rev Bioph Biom. 2003; 32:425–443.

140. Lin H, Truhlar DG. QM/MM: what have we learned, where are we, and where do we go from here? Theor Chem Acc. 2007; 117:185–199.

141. Potestio R, Fritsch S, Espanol P, Delgado-Buscalioni R, Kremer K, Everaers R, Donadio D. Hamiltonian Adaptive Resolution Simulation for Molecular Liquids. Phys Rev Lett. 2013; 110

142. Tozzini V. Multiscale Modeling of Proteins. Accounts Chem Res. 2010; 43:220–230.

143. Zhou H-X. Theoretical frameworks for multiscale modeling and simulation. Curr Opin Struct Biol. 2014; 25:67–76. [PubMed: 24492203]

144. 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]

145. Sherwood P, Brooks BR, Sansom MSP. Multiscale methods for macromolecular simulations. Curr Opin Struc Biol. 2008; 18:630–640.

146. 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]

147. Walsh STR, Cheng H, Bryson JW, Roder H, DeGrado WF. Solution structure and dynamics of a de novo designed three-helix bundle protein. P Natl Acad Sci USA. 1999; 96:5486–5491.

148. Lindorff-Larsen K, Piana S, Dror RO, Shaw DE. How Fast-Folding Proteins Fold. Science. 2011; 334:517–520. [PubMed: 22034434]

149. 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]

150. Zacharias M. Combining coarse-grained nonbonded and atomistic bonded interactions for protein modeling. Proteins. 2013; 81:81–92. [PubMed: 22911567]

151. Fiorucci S, Zacharias M. Binding site prediction and improved scoring during flexible protein-protein docking with ATTRACT. Proteins. 2010; 78:3131–3139. [PubMed: 20715290]

152. 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]

153. 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]

154. 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

155. Dans PD, Zeida A, Machado MR, Pantano S. A Coarse Grained Model for Atomic-Detailed DNA Simulations with Explicit Electrostatics. J Chem Theory Comput. 2010; 6:1711–1725. [PubMed: 26615701]

156. Perez A, Marchan I, Svozil D, Sponer J, Cheatham TE, Laughton CA, Orozco M. Refinenement of the AMBER force field for nucleic acids: Improving the description of alpha/gamma conformers. Biophys J. 2007; 92:3817–3829. [PubMed: 17351000]

157. Machado MR, Dans PD, Pantano S. A hybrid all-atom/coarse grain model for multiscale simulations of DNA. Phys Chem Chem Phys. 2011; 13:18134–18144. [PubMed: 21927740]

158. Villa E, Balaeff A, Mahadevan L, Schulten K. Multiscale method for simulating protein-DNA complexes. Multiscale Model Sim. 2004; 2:527–553.

159. Villa E, Balaeff A, Schulten K. Structural dynamics of the lac repressor-DNA complex revealed by a multiscale simulation. P Natl Acad Sci USA. 2005; 102:6783–6788.

160. Mullinax JW, Noid WG. Recovering physical potentials from a model protein databank. P Natl Acad Sci USA. 2010; 107:19867–19872.

161. Feig M, Sugita Y. Reaching new levels of realism in modeling biological macromolecules in cellular environments. J Mol Graph Model. 2013; 45:144–156. [PubMed: 24036504]

162. Elcock AH. Challenges in Large-Scale Molecular Simulations of Intracellular Environments. Biophys J. 2011; 100:6–6.

163. Thomas AS, Elcock AH. Direct observation of salt effects on molecular interactions through explicit-solvent molecular dynamics simulations: Differential effects on electrostatic and hydrophobic interactions and comparisons to Poisson-Boltzmann theory. J Am Chem Soc. 2006; 128:7796–7806. [PubMed: 16771493]

164. Thomas AS, Elcock AH. Direct Measurement of the Kinetics and Thermodynamics of Association of Hydrophobic Molecules from Molecular Dynamics Simulations. J Phys Chem Lett. 2011; 2:19–24. [PubMed: 26295208]
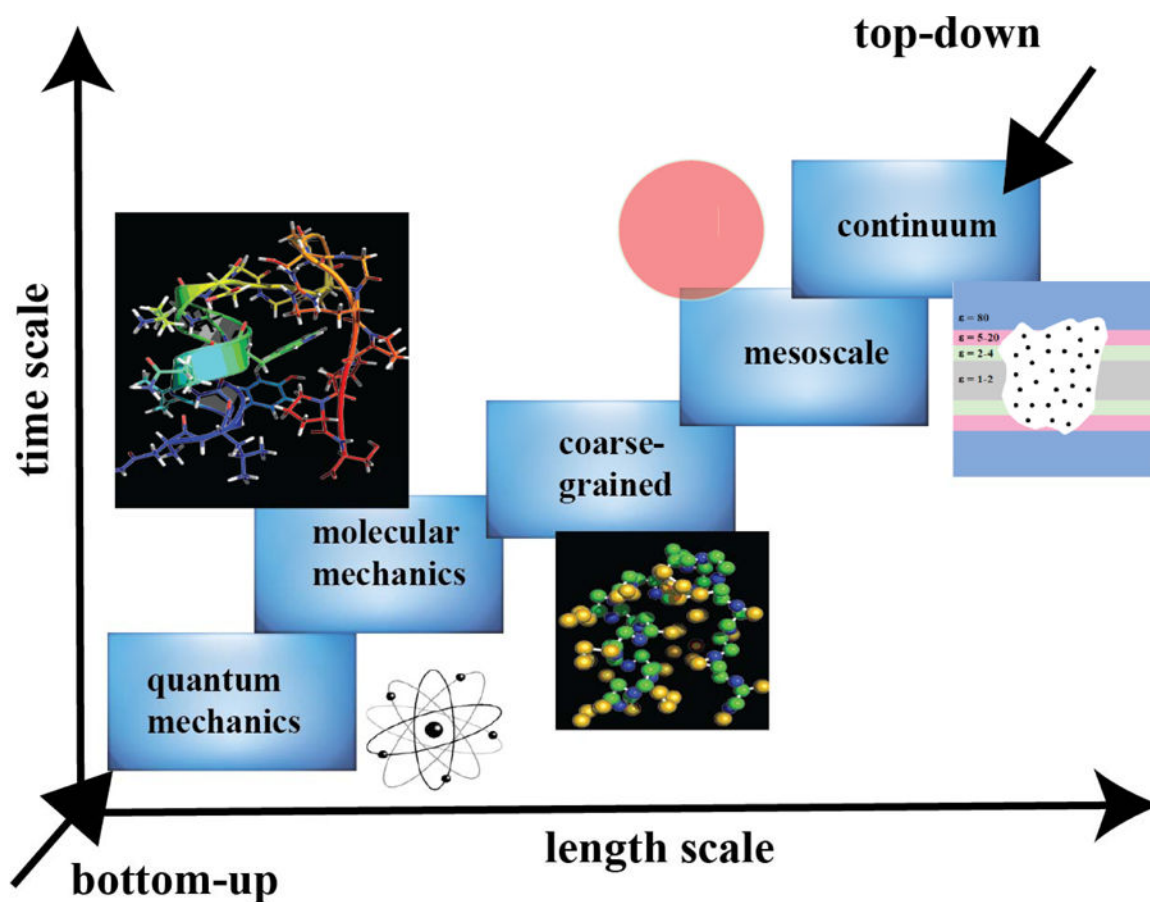
Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1.
Schematic representation of various computational approaches that are used to cover different scales of length and time associated with different biophysical processes. These methods range from highly accurate but also the most computational demanding ab initio molecular dynamics (quantum) in the lower left corner to the highly efficient but very low-detail continuum models. In between, there exist atomistic and coarse-grained models followed by mesoscale models.
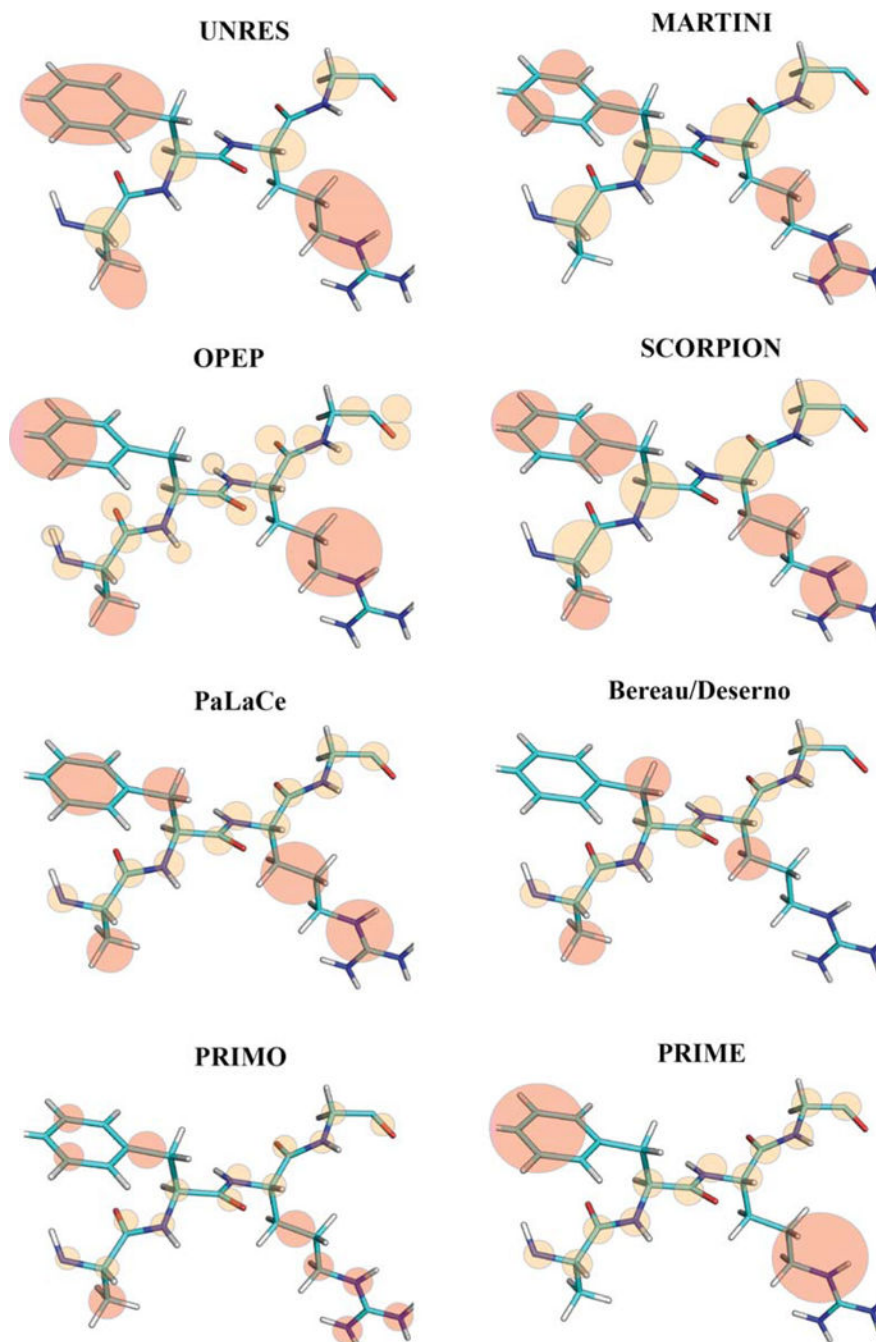
Figure 2.
Mapping strategies for various CG models for an AlaPheArgGly peptide. Backbone beads are shown in yellow color while side-chain beads are shown in red color.
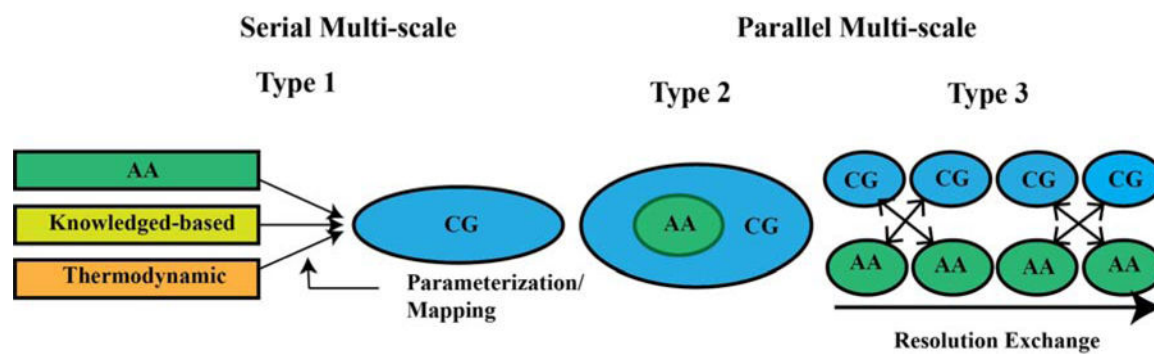
Figure 3.
Schematic of the three different types of multi-scale AA/CG approaches frequently employed in biomolecular simulations.
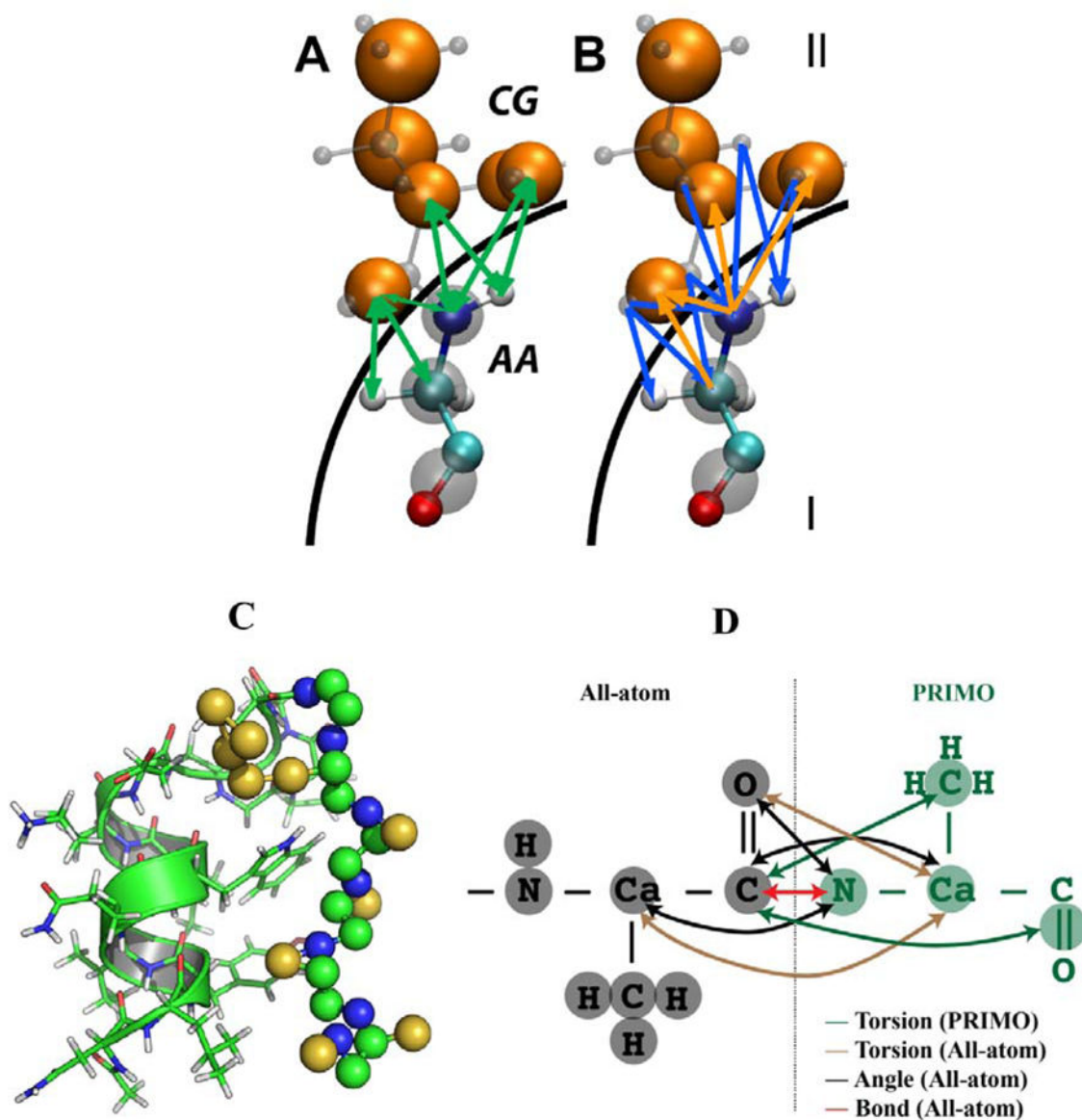
Figure 4.
A schematic representation of coupling of PRIMO and CHARMM force fields when both parts interact with each other via bonded interactions: A, B: coupling schemes; C: Multi-resolution representation of trp-cage where 1–14 residues are described by AA force field (stick); D: interfacial bonded parameters are taken from existing atomistic or coarse-grained force field (black: atomistic, red: atomistic, brown: atomistic, and green: PRIMO).

Table 1

Key features of different popular CG models for proteins. BB denotes backbone while SC denotes side-chain. Here, V^{bond}, V^{angle}, $V^{torsion}$, $V^{double-torsion}$, V^{vdW}, V^{ele}, V^{H-bond}, and $V^{multi-body}$ represent bond, angle, torsion, double-torsion, van der Waals, electrostatic, hydrogen-bonding, and multi-body potentials, respectively.

Model	UNRES	MARTINI	OPEP	SCORPION	PaLaCe	Bereau and Deserno	PRIMO	GBEMP	PRIME
BB	1	1	5	1	3	3	3	2	3
SC (non-Gly)	1	1–5	1	1–2	1–2	1	1–5	2–3	1
Virtual Bead	X	–	–	–	X		X		–
V^{bond}	X	X	X	–	X	X	X	X	X
V^{angle}	X	X	X	–	X	X	X	X	X
$V^{torsion}$	X	X	X	–	X	X	X	X	X
$V^{double-torsion}$	X	–	–	–	–	–	–	–	–
V^{vdW}	G-B(a) 12-6	LJ 12-6	r^{-8}+AF(b)	r^{-8}+G(c)	LJ 8-6	WCA	LJ 12-6	G-B(a) 12-6	Hard-sphere
V^{ele}	X	X	–	X	X	–	X	–	–
V^{H-bond}	–	–	X	–	X	X	X	–	X
Solvent	Effective(d)	CG	Effective(d)	CG	CV(e)	Effective(d)	GBMV(f)	GK(g)	Effective(d)
Membrane	–	X	–	–	–	–	X	–	
Restraint	–	ENM	–	ENM	–	–	–	–	–
$V^{multi-body}$	X	–	–	–	–	–	–	–	–
Dipole	–	–	–	–	–	X	–	–	–
Multipole	–	–	–	–	–	–	–	X	–
Speedup	~1000	75–100	30–40	100–130	~1000	NA(h)	10–15	~50	~1000

(a) Gay-Bern,

(b) r^{-8} combine with analytical formulation,

(c) r^{-8} combine with Gaussian attractive,

(d) implicitly included within nonbonded parameters,

(e) circular variance,

Author Manuscript

(f)Generalized Born/Molecular Volume,

(g)Generalized Kirkwood,

(h)Not available.

Author Manuscript

Author Manuscript

Author Manuscript