CHEMICAL REVIEWS

Review

pubs.acs.org/CR

Coarse-Grained Protein Models and Their Applications

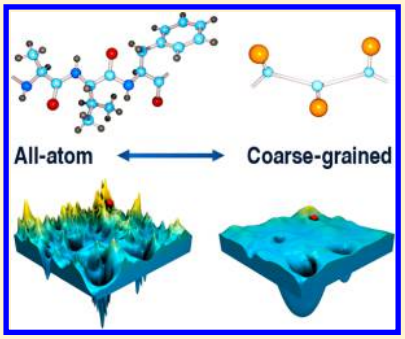
Sebastian Kmiecik,[†] Dominik Gront,[†] Michal Kolinski,[‡] Lukasz Wieteska,[†,§] Aleksandra Elzbieta Dawid,[†] and Andrzej Kolinski*[,†]

[†]Faculty of Chemistry, University of Warsaw, Pasteura 1, 02-093 Warsaw, Poland

[‡]Bioinformatics Laboratory, Mossakowski Medical Research Center of the Polish Academy of Sciences, Pawinskiego 5, 02-106 Warsaw, Poland

[§]Department of Medical Biochemistry, Medical University of Lodz, Mazowiecka 6/8, 92-215 Lodz, Poland

ABSTRACT: The traditional computational modeling of protein structure, dynamics, and interactions remains difficult for many protein systems. It is mostly due to the size of protein conformational spaces and required simulation time scales that are still too large to be studied in atomistic detail. Lowering the level of protein representation from all-atom to coarse-grained opens up new possibilities for studying protein systems. In this review we provide an overview of coarse-grained models focusing on their design, including choices of representation, models of energy functions, sampling of conformational space, and applications in the modeling of protein structure, dynamics, and interactions. A more detailed description is given for applications of coarse-grained models suitable for efficient combinations with all-atom simulations in multiscale modeling strategies.



CONTENTS

1. Introduction 7898
2. Coarse-Grained Protein Models 7900
2.1. Brief History 7900
2.2. Levels of Resolution 7901
2.3. Force Fields 7901
2.3.1. Physics-Based Force Fields 7902
2.3.2. Knowledge-Based Statistical Force Fields 7904
2.3.3. Structure-Based Models of Force Field 7904
2.4. Sampling Schemes 7905
2.5. Examples of Protein Coarse-Grained Models 7907
3. Multiscale Modeling 7908
3.1. Example Strategies 7908
3.2. Reconstruction of Atomic Representation from Coarse-Grained Models 7909
4. Applications 7911
4.1. Coarse-Grained Models in Multiscale Modeling Pipelines 7911
4.2. Simulations of Protein Dynamics 7911
4.2.1. Protein Folding and Large Scale Dynamics 7911
4.2.2. Protein Flexibility: Small-Scale Dynamics 7914
4.3. Protein Structure Prediction 7914
4.3.1. Importance of Computational Structure Prediction 7914
4.3.2. Comparative Modeling 7915
4.3.3. Ab Initio Modeling 7915
4.3.4. Critical Assessment of Protein Structure Prediction Methods 7915
4.4. Protein Interactions 7916
4.5. Membrane Proteins 7919
4.6. Integrative Modeling 7920
5. Concluding Remarks 7920
Author Information 7921
Corresponding Author 7921
Notes 7921
Biographies 7921
Acknowledgments 7922
References 7922

1. INTRODUCTION

Living organisms are the most complex chemical systems whose function depends on a vast number of molecules, from simple monomers through many small and medium-size oligomers and copolymers (peptides, proteins, RNA, etc.) to huge copolymers such as DNA. With some exceptions, proteins are composed of 20 types of amino acids. Again with some exceptions, all amino acids in living organisms have left-handed conformations. Since typical protein chains consist of a few tens to hundreds of amino acids, the number of possible amino acid sequences of such copolymers is enormous. While sequences of amino acid units in natural proteins look at first glance random, they are certainly not.[1] The majority of known natural proteins fold into specific three-dimensional structures, while the vast majority of random polypeptides collapse to somewhat less dense unstructured states. Protein folding plays an essential functional role in living cells, although this process could be also observed at properly controlled in vitro experiments. Owing to the impressive progress in the experimental methods

Received: March 1, 2016
Published: June 22, 2016

DOI: 10.1021/acs.chemrev.6b00163
Chem. Rev. 2016, 116, 7898−7936

of molecular biology in the last decades we now know around 120 thousand three-dimensional native-like protein structures or their complexes, with resolutions from about 0.5 to 2−3 Å. This is still only a small fraction of proteins with known sequences,[2] although for a large fraction of sequenced proteins their three-dimensional structures can be predicted theoretically by various combinations of bioinformatics and molecular modeling techniques.[3] Theoretical prediction of folded (native-like) three-dimensional protein structures is just one of the key tasks of computational structural biology. Native structures are not completely fixed,[4−6] but they change when proteins perform their biological function, interact with other biomacromolecules, or undergo unfolding−folding transitions. Computational modeling of these processes is crucial for creating realistic molecular pictures of biological protein functions, interpretation of different experimental data, knowledge-based drug design and various aspects of biotechnology, etc.[7−13] Classical atom-level molecular modeling can address many of the these tasks, but its practical applications are still limited by its algorithmic efficiency and the available computing power.[14] Even using a special-purpose supercomputer dedicated to atomistic molecular dynamics (MD) simulations,[15] it is possible to simulate folding processes of only small, relatively fast folding proteins[16,17] or their dimerization processes[18] (see Figure 1). Similar limitations apply to molecular docking,

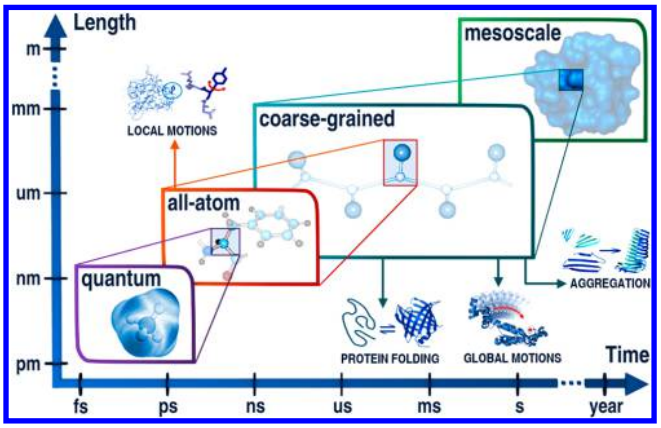


Figure 1. Application ranges for molecular modeling at different resolutions: quantum, all-atom, coarse-grained, and mesoscale. The plot shows approximate ranges of time scales and system sizes (lengths). The presented application ranges can be expanded by merging tools of different resolution into multiscale schemes.

studies of dynamics of biomacromolecular systems, and other related tasks. This is a major reason why development and practical applications of coarse-grained protein modeling methods is needed.[19−33]

The first coarse-grained protein models were proposed almost half a century ago, although only very recently did coarse-grained models become widely used, especially in multiscale modeling pipelines. In 2013, the Nobel Prize Committee awarded the Prize in Chemistry "for the development of multiscale models for complex chemical systems," recognizing the early achievements of Michael Levitt, Ariel Warshel, and Martin Karplus that included the coarse-grained modeling of proteins[34,35] as an important step in the investigation of large biomolecular systems.[36]

In the last ten years the number of publications on developments and applications of coarse-grained models of biomolecules has increased several times.[27] There are good reasons for this increasing role. First, experimental molecular biology provides enormous volumes of data that need interpretation. Second, as noted before, in spite of the rapid increase of computing power, applications of all-atom MD, the classical tool for molecular modeling, are still limited to relatively small systems and rather fast processes. Coarse-grained models are computationally more effective and enable simulations of much longer time-scales and/or larger sizes of the systems studied. Third, well-designed coarse-grained models of a not too low resolution enable reasonable reconstruction of modeled structures to all-atom resolution. This opens up a possibility of multiscale modeling, based on a combination of the computational speed of coarse-grained models with the high accuracy of classical all-atom MD.[25,37−40]

Coarse-grained protein models assume various levels of reduced polypeptide chain representation[19−32] (see section 2.2). The protein main chain could be represented by all heavy atoms or by one or two united atoms per residue, while just one or two united atoms typically replace the side chain. Various definitions of models of interactions for coarse-grained representations are possible (see section 2.3). Perhaps more challenging "physics-based" derivations of coarse-grained force fields start from classical all-atom models of interactions and translate them into united atom potentials.[21,30] Very different are "knowledge-based" interaction schemes derived from the statistical regularities seen in known protein structures.[41,42] Both approaches to building interaction schemes have their weaknesses and advantages. Sampling procedures can be based on various versions of MD and/or Monte Carlo (MC) methods (see section 2.4). Sometimes heuristic approaches are also used. The majority of coarse-grained models use continuous representations of the geometry of modeled structures. Few coarse-grained models use lattice grids which enable significant computational speedup compared to continuous models.[43,44] Obviously, to achieve good resolution of a model the lattice needs to be of dense spacing, which enables high coordination numbers for the location of neighboring united atoms.

Many useful applications of coarse-grained protein models have been described in the past few years.[19−32] Coarse-grained models have been successfully used in studying protein folding mechanisms based on either very generalized protein-like models or simulations of real proteins (see section 4.2). Another productive area for coarse-grained modeling is protein structure prediction. Every two years CASP (Critical Assessment of Protein Structure Prediction) experiments provide a good test of computational methods applied for structure prediction (see section 4.3.4). Most leading groups successfully use coarse-grained modeling tools, which are the methods of choice in the most difficult de novo modeling cases, although coarse-grained simulations also play a significant role in the advanced tools of comparative (homology based) modeling.[3,45−48]

This review focuses on the prospective applications of coarse-grained models, including protein structure prediction, modeling of complex dynamic processes, protein interactions with other proteins and peptides and modeling of membrane proteins. Some of the most successful applications have been recently achieved by the combination of coarse-grained models with a wide range of computational techniques, including classical all-atom modeling and careful implementation of restraints derived from various sources of experimental data.

DOI: 10.1021/acs.chemrev.6b00163
Chem. Rev. 2016, 116, 7898−7936

Future developments are expected to continue on this integrative modeling trend.[33,49−55]

2. COARSE-GRAINED PROTEIN MODELS

2.1. Brief History

It has been about half a century since we have been witnessing rapid progress of experimental structural biology. In particular, we learned that proteins adopt specific three-dimensional structures, essential for most of their biological functions. At the same time, due to the rapidly increasing computer power and progress in theoretical chemistry and physics, it became clear that molecular modeling of biomacromolecules could be essential for understanding molecular backgrounds of many biological processes. It also became clear some time ago, and at different levels it remains true today, that the productive applicability of classical atom-level molecular modeling of biological systems is limited to relatively small systems and biologically short time scales (see Figure 1). In this context it was crucial to solve the so-called Levinthal paradox.[56] It was known that globular proteins, or at least a large fraction of them, adopt well-defined three-dimensional structures. According to Levinthal, such a process, with a random search for all possible conformations of a protein chain consisting of one hundred residues, would take longer than the age of the universe. Therefore, the folding process could not be fully random, since proteins of such size form their unique structures usually in a time range of milliseconds. Today, thanks to many theoretical and experimental studies, it is well understood that local secondary structure preferences and other geometric features of protein chains drastically decrease the number of available conformations and therefore facilitate relatively fast folding to native tertiary structures.

Understanding the protein folding mechanisms was probably one of the main reasons for designing coarse-grained protein models to perform simulations at time scales important for biological processes. The first coarse-grained protein models were developed almost 40 years ago. The classical work of Levitt and Warshel[34] already cited in the introduction is a good example of successful early attempts at simulating an entire folding process. In their protein model, a chain of pseudoatoms (placed at Cα positions) replaced the main chain structure. United pseudoatoms at centers of their average conformations (see Figure 2) replaced the side chains, except for glycine. The planar angle between three consecutive Cα was assumed constant, equal to the statistical average seen in the known protein structures. This was crude simplification due to the significant difference between the average values of these angles observed in various secondary structure fragments. Positions of three consecutive Cα atoms defined the center position of the side chain for the second residue. The only degree of freedom in this model described rotation along the central pseudobond for the three consecutive Cα atoms. A simple Lennard-Jones potential described interactions between the united atoms. Brownian dynamics (BD) was used as the sampling scheme. A large series of BD simulations were performed for the bovine pancreatic trypsin inhibitor, a small protein, and in some runs native-like low resolution structure models were obtained. This work clearly demonstrated that the packing and pairwise interactions of side chains are one of the main forces leading to specific folded structures. A year later Levitt[57] proposed a slightly more accurate version of this model, accounting for the variable orientation of the united side chains. The torsional

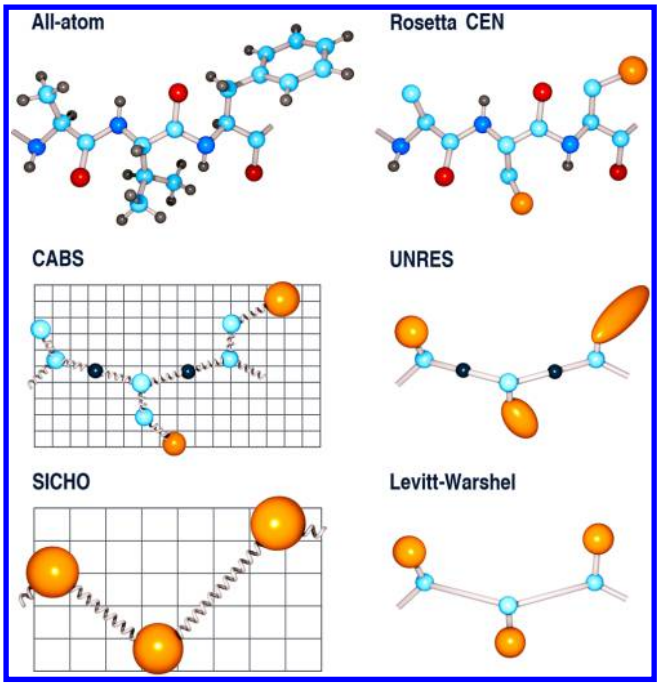


Figure 2. All-atom representation of a tripeptide and the corresponding coarse-grained models. Various coarse-grained models are presented: Rosetta centroid mode (CEN) representation,[71] CABS,[72] UNRES,[73] SICHO,[74] and Levitt and Warshel model.[34] United side chain atoms are colored in orange. Pseudobonds of fluctuating length are shown as springs and lattice models are shown on the underlying lattice slide.

potential for the main chain (and a side chain, where applicable) degrees of freedom was developed from the statistical analysis of conformational properties of representative dipeptides. Similar models were studied by Hagler and Honig[58] and by Wilson and Doniach.[59] In the latter study the authors proposed interesting derivation of statistical potentials for residue−residue interactions and used the Monte Carlo method for simulated annealing simulations of the folding process. An interesting design of the side chain empirical potential was also proposed by Crippen and Ponnuswamy.[60] Related reduced models of small proteins, with various approaches to the force field problem and sampling strategies, including continuous and lattice representations, were published by many authors,[19−33,43,44,61−70] and we cite only selected early reports.

A different area of research focused on protein-like models.[75] Dill,[76] Shakhnovich,[77] Chan,[78] and many others studied simple cubic lattice chains and their folding to unique three-dimensional structures.[78,79] Most of these models treated protein chains like sequences of two types of amino acids: hydrophobic (H) and polar (P), although somewhat more detailed models were also investigated.[80] It was clearly demonstrated (with exact statistics of all possible conformations of HP models) that specific sequences are needed for unique structures, and a rigorous description was formulated of dynamics and thermodynamics of folding processes in these idealized systems. Intermediate resolution lattice models (between idealistic "protein-like" and crude protein models) such as diamond lattice[81] or "chess-knight"[82,83] models were also studied. As pointed out by Park and Levitt[84] and Godzik et al.,[85] while very attractive for their simplicity and allowing exact statistical analysis, the low resolution lattice models of proteins

DOI: 10.1021/acs.chemrev.6b00163
Chem. Rev. 2016, 116, 7898−7936

could also be strongly biased due to their crude representation of mutual orientations of protein chain fragments. Nevertheless, these and other studies on low resolution models provided a strong foundation for the development of contemporary medium and high resolution coarse-grained models that we focus on in this review.

We hope that this short overview of the earlier studies of simplified protein models briefly explains various avenues that led to the present state of the art in coarse-grained protein modeling. In the main sections of this review we describe the recent advances in coarse-grained modeling, focusing on the models which are not only computationally attractive but also give realistic reconstruction of atom-level views of the structure and dynamics of protein systems. We also discuss promising applications of coarse-grained protein models in the multiscale modeling of large biomacromolecular systems.

2.2. Levels of Resolution

Proteins are particular semiflexible oligomers that form specific linear sequences of amino acids. Amino acids are linked by covalent peptide bonds, almost always adopting relatively rigid trans conformation. This imposes some limitations on the conformational space available for the protein backbone chain. With a couple of exceptions, orientations of the side chains of amino acids are asymmetric, and proteins of living organisms usually adopt the L-handed conformation. The side chains are of different size and geometry defined by their internal degrees of freedom and interactions with their environment. This results in additional biases in the polypeptide conformational space.

Characteristic three-dimensional protein structures are determined not only by the conformational properties of the main chain, but also by the resulting specific packing and interactions of the side chains. Interestingly, many (sometimes multiple) random mutations do not change the primary properties of natural proteins, while a different single mutation can not only destroy their biological functions, but also their structural properties.[86] In other words, natural proteins are very specific copolymers "edited" by evolution and the well-defined three-dimensional structures of many of them result from a very complex interplay of main chain flexibility, patterns of hydrogen bonds and interactions between amino acid side chains. Therefore, coarse-grained models, their representation of protein chains, force fields, and sampling techniques must be carefully designed, with the purposes and expected reliability of such models taken into deep consideration.[87]

A broad spectrum of coarse-grained protein chain representations were discussed in literature.[19−33] In all cases, the main purpose was to reduce the number of degrees of freedom treated in an explicit fashion. For this reason, pseudoatoms replace amino-acid fragments or even entire amino acids (an equivalent term used in literature for "pseudo-atom" is "united atom"). The conformational space of these reduced atoms could be also restricted, leading to additional reduction of the available degrees of freedom. The simple lattice protein-like HP models mentioned in the introduction are the extreme examples of this kind of simplification. In an HP cubic-lattice model[75] the two types of amino acids (H-hydrophobic and P-polar) are restricted to a single lattice position. Thereby the mutual orientations of all residues are also restricted to cubic lattice angles. The conformational space of such models (for chains of a limited length) could be exactly enumerated and the behavior of various sequences of HP units precisely studied. Although this level of simplification ignores many important features of real proteins, the studies of HP (and related) models explained some fundamental features of protein-like polymers.[75,78] It is also possible to consider even lower resolution models that enable crude representation of huge protein systems.[88−94]

In contrast to simple HP models it is possible to design a structurally more realistic model based just on a single pseudoatom per amino acid residue. A practical example is the SICHO (side chain only) model in which only explicitly simulated pseudoatoms are placed near the centers of the amino acid side chains[74,95] (see Figure 2). The excluded volume, distribution of distances between consecutive pseudoatoms, distribution of planar angles (dependent on amino acid identity), and torsional angles used in the model were derived from the statistical analysis of structural regularities observed in known protein structures. These distance and angle restrictions were additionally controlled by the crude approximation of main chain geometry automatically fitted to the positions of the side chain pseudoatoms of three consecutive amino acids. The SICHO resolution is about 2−3 Å and provides a crude although quite realistic representation of protein structure, including an acceptable picture of secondary structure.[96,97] Since the side chains in proteins are most mobile and their contacts are crucial for the packing of protein structures, the modeling schemes led by SICHO pseudoatoms motion enable extremely efficient simulations of protein dynamics, especially in the dense (near-native) state. Unfortunately, this intermediate resolution coarse-grained model was not extensively studied, in spite of quite promising preliminary results.[98−100]

The first intermediate resolution coarse-grained models were developed a long time ago, providing much deeper understanding of protein physics and defining new directions in the development of novel methods of the multiscale modeling of proteins and other biomacromolecules.[19−32,37−40] The efforts of Levitt and co-workers initiated this direction,[34,35] see section 2.1. Typical intermediate resolution models use one or two united atoms to approximate the geometry of the main chain and side chains, respectively. Two examples of such models, which allow a very broad range of applications, from structure prediction to a study of protein dynamics and interactions, are UNRES (united residue)[73] and CABS (C-alpha, beta, and side chain)[72] models (see Figure 2 and section 2.5 for their discussion). UNRES and CABS enable quite realistic, although by no means exact, reconstruction to atomistic models.[101−103]

Almost exact coarse-grained protein models, like Rosetta[71,104] or PRIMO,[105,106] treat protein chain representation closely to the atomistic level by introducing only some small simplifications to speed up simulations but also for use in high-resolution modeling. Various versions of Rosetta models use interesting combinations of realistic coarse-grained resolution (see Figure 2) and all-atom representation designed specifically for efficient structure prediction and not for studying folding dynamics. For further details and examples of other coarse-grained representations, see section 2.5.

2.3. Force Fields

Designing force fields for coarse-grained models is to some extent directed by the chosen level of resolution and the expected ranges of applicability, although it is also the result of a different philosophy of thinking about the molecular picture of biological systems. On one hand, there are efforts to build

DOI: 10.1021/acs.chemrev.6b00163
Chem. Rev. 2016, 116, 7898−7936

classical coarse-grained force fields based on molecular physics. On the other hand, the force fields of coarse-grained models could be derived from the statistical analysis of structural (and dynamic) regularities seen in the growing databases of experimental structures, underestimating the atomic-level backgrounds. Various combinations of these two fundamentally different approaches are possible.[107] Typically, in comparison to its all-atom counterpart, the coarse-grained force field smoothens out the energy landscape, and thereby helps to avoid local energy minima "traps," see Figure 3. Coarse-graining

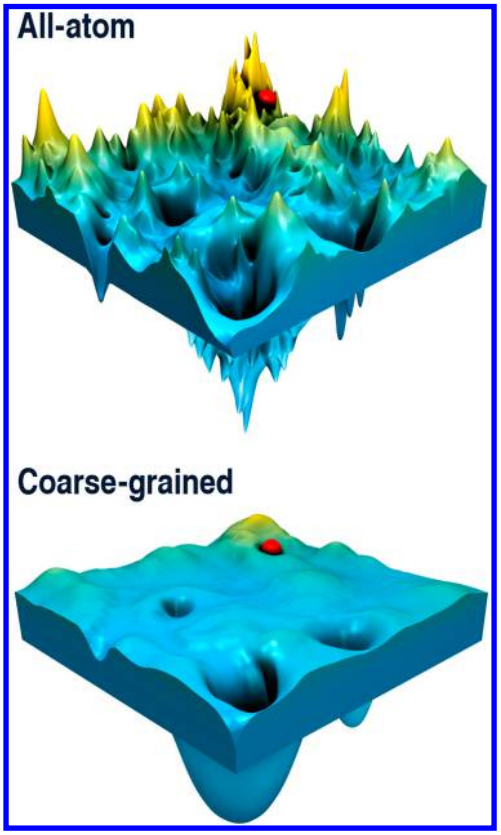


Figure 3. All-atom versus coarse-grained energy landscape. The figure illustrates the effect of the smoothening of the energy landscape in a coarse-grained model as compared to an all-atom model. The flattening enables efficient exploration of the energy landscape in search for the global minima, while avoiding traps in the local minima.

also affects thermodynamic properties of a modeled system, particularly the balance between enthalpy and entropy. Reduction of the degrees of freedom affects the entropy of the simulation system, which is compensated by reduced enthalpic terms. In turn, a coarse-grained model may accurately reproduce free energy differences but contributing enthalpy and entropy values may be inaccurate. Such limitations are typical for the majority of coarse-grained models.

2.3.1. Physics-Based Force Fields. A general formula for a classical physics-based all-atom force field consist of six terms[108](eq 1). The first four of them, so-called "bonded" terms, describe bonds deformation (eq 1a), bond angles geometry (eq 1b), and rotation about certain dihedral angles (eq 1c and 1d). The last two "non-bonded" terms describe dispersion and repulsion effects (Lennard-Jones term, eq 1e) and electrostatic interactions (Coulombic term, eq 1f).

$$U(\vec{R}) = \sum_{\text{bonds}} K_b(b - b_0)^2 \tag{1a}$$

$$+ \sum_{\text{angles}} K_\theta(\theta - \theta_0)^2 \tag{1b}$$

$$+ \sum_{\text{dihedrals}} K_\chi(1 + \cos(n\chi - \delta)) \tag{1c}$$

$$+ \sum_{\substack{\text{improper}\\ \text{dihedrals}}} K_{\text{imp}}(\varphi - \varphi_0)^2 \tag{1d}$$

$$+ \sum_{\substack{\text{nonbonded}\\ \text{pairs } i,j}} \left(\varepsilon_{ij} \left[\left(\frac{R_{\min_{ij}}}{r_{ij}} \right)^{12} - \left(\frac{R_{\min_{ij}}}{r_{ij}} \right)^{6} \right] \right) \tag{1e}$$

$$+ \sum_{\substack{\text{nonbonded}\\ \text{pairs } i,j}} \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon r_{ij}} \tag{1f}$$

The equation above looks relatively simple but we have to keep it in mind that the summations involve many different atom types and, therefore, many parameters have to be defined. These values are defined using quantum-mechanical calculations or experimental measurements (hence the name "experimental force field"). Bond lengths with corresponding stiffness values (b_0 and K_b in eq 1a) as well as bond angle parameters (θ_0 and K_θ in eq 1b) are usually obtained from crystallographic and spectroscopic data for small molecules. Parameters for the Lennard-Jones term are usually optimized using data from small molecule liquid density, heat of evaporation, or free energies of solvation.[109] Partial atomic charges (q_k in eq 1f), necessary for the evaluation of the Coulombic term, are obtained from QM calculations. Therefore, from a certain perspective, these force fields may be seen already as coarse-grained with electronic degrees of freedom averaged out.

In general, a physical-based coarse-grained force field can be described by a similar formula as an all-atom force field (eq 1a−1f). In practice, a broad variety of additional expressions going beyond the classical formula are used to describe the energy of coarse-grained models. During the coarse-graining process some atoms are removed and the degrees of freedom related to them are averaged out. In this situation, internal correlations between groups of atoms (now represented as united atoms) must be introduced explicitly in the form of multibody terms. Most approaches keep the distinction between local energy terms and so-called contact potentials. The former describe spatial correlations between pseudobond vectors which no longer follow classical laws such as a harmonic function and are often expressed by a kind of an arbitrarily chosen function (like Chebyshev polynomials,[110] splines, or histograms). Nonbonded terms are usually represented by a single formula depending on the type of interacting atoms, the distance between them and, sometimes, their mutual orientation and local neighborhood. Due to the diffused nature of a spherical cloud representing a group of physical atoms, these interactions tend to be softer and the 12-6 exponents in the van der Waals equation may be substituted with more appropriate numbers.[111] How the final formulas of the coarse-grained force fields look depends not only on the specific

DOI: 10.1021/acs.chemrev.6b00163
Chem. Rev. 2016, 116, 7898−7936

model of coarse-graining but also on the chosen method of transferring atomistic formulas onto coarse-grained (united atoms) potentials; good examples can be found in recent reviews.[30,112] Below we provide an overview of different approaches for derivation of physics-based different coarse-grained force field terms.

The least invasive step in coarse-graining is to neglect nonpolar hydrogen atoms. This yields up to 10-fold reduction of computational effort. CHARMM19[113,114] and GROMOS[115] are examples of such type of a commonly used force field. Parametrizations of these force fields were done in the same way as described above for the classical all-atom case. When further reducing a molecular representation it becomes necessary to determine physical parameters (as radii) for unphysical moieties (coarse-grained pseudo atoms). A remarkable example is the MARTINI force field,[116] which has been parametrized by reproducing the partitioning of free energies between polar and apolar phases of a large number of chemical compounds (see section 2.5).

The strategy to derive force field parameters from experimental data becomes increasingly difficult with an increasing number of distinct pseudoatom types in coarse-grained representation. Below we outline a few systematic approaches, proposed in the literature, to derive a coarse-grained force field from results of all atom simulations of a system (or systems) that were conducted with a reference ("true") all-atom potential $u^{AA}(r)$. Parameters for the corresponding coarse-grained force field are derived to match some features of the atomistic ensemble. The latter are usually collected from all-atom molecular dynamics simulations but quantum-mechanical approaches have also been exploited.[117,118]

In the iterative Boltzmann inversion (IBI) approach, which was first proposed by Schommers,[119] radial distribution function (RDF) or functions $\rho^{AA}(r)$ are used as the target property. Starting from an initial form of a coarse-grained pair potential $u^{CG}(r)$, a coarse-grained RDF is calculated and used to improve the estimation of $u^{CG}(r)$ according to the formula:

$$u_{i+1}^{CG}(r) \cong u_i^{CG}(r) - k_B T \ln\left(\frac{\rho_i^{CG}(r)}{\rho^{target}(r)}\right) \quad (2)$$

where T is the absolute temperature and subscript indexes denote iteration of the Boltzmann inversion procedure.

This procedure is repeated until successful convergence is achieved, which, according to the Henderson theorem,[120] leads to the pair potential that is unique for a given $\rho(r)$. In practice however many pair potentials are able to reproduce a target $\rho(r)$ within an acceptable error. Reith et al.[121] applied the IBI procedure for two model systems of a known Hamiltonian (Weeks–Chandler–Andersen and Lennard-Jones potentials[121]) and concluded that, even though the procedure results in an RDF that is undistinguishable from the target, the resulting potential still differs from the true one. A possible remedy might be to use additional target properties that may be included in the IBI calculations,[122] such as pressure.[121] Nevertheless the IBI framework has become very popular due to its simplicity and quick convergence, and usually several iterations are required. In the most simplistic application of this method, only one iteration of Boltzmann inversion is conducted.

An all-atom RDF was also chosen as the target property in the original formulation of the inverse Monte Carlo (IMC) method[123] for deriving a coarse-grained Hamiltonian defined as a linear combination of terms. Similarly to IBI, initial approximation to the coarse-grained Hamiltonian is iteratively improved to minimize the difference between all-atom reference simulations and coarse-grained RDFs. When compared to IBI, the IMC method explicitly handles correlations between coarse-grained force field parameters. Coarse-grained force field parameters are calculated in an iterative process where at each step a set of linear equations is solved to find a better approximation to FF parameters. The subsequent generalization of the IMC method known as Newton inversion[124] may utilize virtually any property derived from all-atom simulation as the target distribution and the mathematical formulation of the coarse-grained energy function is no longer restricted to any particular form. The method uses the Newton–Raphson approach to iteratively solve a set of nonlinear equations.

The methods described above attempt to preserve the RDF as much as possible during coarse-graining. In turn, the force matching approach[118] derives pairwise forces acting on coarse-grained sites to match atomistic forces calculated for a set of reference conformations. In the seminal formulation of FM, ab initio calculations were used as the "true" potential. Cubic spline was used to represent the coarse-grained forces and the necessary spline parameters were fit by minimizing the mean-square error between coarse-grained and atomistic forces averaged over all trial configurations. The method was further developed by Voth[117,125] (under the name of multiscale coarse-graining or MS-CG) who used MD trajectories as the reference. To improve the accuracy of resulting coarse-grained potentials, an alternative iterative scheme of force matching has been recently proposed.[126,127] Convergence of the iterative procedure was reported as much faster than for the IBI approach (both IBI and iterative FM[128] may be derived on the grounds of Yvon-Born–Green theory).

Relative entropy minimization (REM) relies on minimizing the Kullback–Leibler divergence (relative entropy) between an all-atom and a coarse-grained system.[129] This parameter measures the degree of overlap between all-atom and coarse-grained distributions of states and has non-negative values with zero meaning a perfect overlap. Therefore, unlike IMC and IBI, the REM method in general may be used to quantitatively compare different coarse-grained schemes or different mathematical forms of a coarse-grained Hamiltonian. The actual minimization of relative entropy may be done by the steepest descent or Newton–Raphson[129] or stochastic[130] minimization. An even more efficient algorithm that uses statistics sampled from coarse-grained trajectories was recently proposed by Shell.[131] The formulation of the REM method is very general and may be applied to optimize coarse-grained structure mappings[132] and dynamics.[133] Correspondence between REM and other methods reported above is discussed in a couple of interesting publications.[131,134]

Conditional reversible work (CRW),[135] yet another approach to coarse-grained force field parametrization, relies on a thermodynamic cycle to calculate energy of two coarse-grained sites. The energy (reversible work) is calculated between groups of atoms in their natural chemical environment when restrictions are imposed on the mutual orientations that can be adopted by these two groups due to surrounding chemical moieties. The CRW method was recently extended to derive dissipative particle dynamics friction functions.[136]

DOI: 10.1021/acs.chemrev.6b00163
Chem. Rev. 2016, 116, 7898–7936

The coarse-graining approaches outlined above have been extensively used in various studies related to biomolecules in their natural environment. This most notably includes lipid bilayers and water. Coarse-grained approaches to membranes are described in section 4.5. Water coarse-grained models were recently reviewed elsewhere.[137] Despite the differences between the coarse-grained strategies described above, they all require a reference ("true") ensemble of all-atom conformations. It is important to keep it in mind that the reference data were collected in particular conditions such as temperature, concentration or even size of the simulated system. These variables are implicitly incorporated into the resulting coarse-grained force field which in general may be not applicable to other conditions. The problem of transferability between different environments requires careful analysis.[30]

2.3.2. Knowledge-Based Statistical Force Fields. Designing effective transferable force fields for coarse-grained representations based on atom-level potentials is a challenging task, as outlined in the previous section. On the other hand, we know a huge number of experimentally determined protein structures. The details of conformational features and atomic packing, controlled by complex interactions, may be analyzed on the grounds of statistical analysis. This led to the idea of knowledge-based statistical potentials.

In their seminal work Tanaka and Scheraga[138] derived a coarse-grained potential based on their study of relative frequency of atomic contacts observed in the crystal structures of proteins. Interaction between two types of amino acid side chains was defined as

$$\frac{-E}{k_B T} = \ln\left(\frac{N_{\text{observed}}}{N_{\text{reference}}}\right) + c \tag{3}$$

where N_{observed} is the observed frequency of contacts of specific side chains and $N_{\text{reference}}$ is the expected frequency observed in the reference state.

The procedure clearly resembles Iterative Boltzmann Inversion. Unlike IBI, however, the derivation of knowledge-based potentials is usually a single-step process. Thus, the choice of the reference state, which is a priori unknown, is crucial for this approach. Knowledge-based potentials can be also derived using the Bayes theorem. Following the RAPDF (residue-specific all-atom conditional probability discriminatory function) potential formulation given by Samudrala and Moult,[139] we can define energy as a logarithm of the probability that a particular conformation is "correct," given its amino acid sequence A and a vector of structural features X (such as distances, contacts, angles, etc.):

$$E = \ln(P(\text{C}|X, A) = \ln\left(\frac{P(X, A|C)}{P(X, A)}\right) + c \tag{4}$$

where $P(X,A|C)$ is the distribution observed in the population of "correct" structures. Most commonly a nonredundant subset of PDB deposits is used for this purpose. The reference state should be interpreted as the a priori distribution $P(X,A)$. A very similar formalism was applied by Baker and co-workers[140] to derive the Rosetta force field. The correspondence of Boltzmann and Bayesian approaches has been described by Xia and Levitt.[141]

The hypothetical reference states should be of the same volume (and shape) as the real (observed) structure. We could use, for instance, a set of similar protein structures with a random sequence of amino acids, but with the same composition. Miyazava and Jernigan[61] introduced the concept of quasi-chemical approximation: a perfect mixture of spheres representing the 20 amino acid types with their molar fractions corresponding to the probability of finding such an amino acid in a database. That concept was further formalized by Sippl.[142] In practice these molar fractions should account for the finite size of a protein sequence, and thus proper composition corrections are required.[143−146] All-atom knowledge-based contact potentials may be derived in the same manner[147] as in the case of DOPE,[148] dDFIRE,[149] GOAP,[150] or ROTAS[151] energy functions. In a similar way we can also define statistical potentials describing short-range interactions,[152] angular preferences[153] geometric aspects of hydrogen bonding,[154] etc.

Apart from those mentioned above, alternative methods for the derivation of knowledge-based force fields were also proposed based on direct optimization of force field performance. Various criteria of success were proposed, including: maximizing the energy gap between the "native" and "non-native" conformation,[155] maximizing the native energy z-score,[156,157] maximizing the probability of successful prediction[158] or minimizing the free energy of the native state.[159] Several formalisms, interpretations, and extensions of knowledge-based force fields have been published[61,140,141,143,160−165] and recently reviewed.[107,166,167]

While the basic schemes of knowledge-based interaction models generally follow the structure of physic-based force fields, as defined by eq 1, their derivation can be conceptually more challenging. Depending on the level of coarse-grained representation, definition of the model force field, and the complexity of the experimental databases used, the final formulas may be composed from a significantly larger number of specific terms than that given in eq 1. Moreover, some terms of a knowledge-based force field can describe specific conditional combinations of bonds, angular and nonbonded interactions. One of the most extreme applications of the knowledge-based strategy is probably the statistical force field designed for the CABS model of single domain globular proteins.[72] The CABS force field treats amino acid interactions in a context-dependent way and takes into account very complex multibody effects, encoded in a large number of composition and structure encoding parameters (for details of CABS, see section 2.5). CABS and other coarse-grained models[168−170] based on context-dependent potentials, such as CAS (implemented in I-TASSER method[48]), are now one of the most effective tools in de novo structure prediction[171−173] (see also section 4.3).

The strength of knowledge-based force fields emerges from their simplicity and efficiency in protein structure prediction, modeling of protein folding pathways, and related tasks of computational biology. The weak point is a lack of transferability. While the force field derived for single domain globular proteins will work well for a vast majority of single protein and peptide structures, the interaction between independent domains, interactions between proteins and nucleic acids, etc. require derivation of a new component of knowledge-based force fields. Fortunately, rapidly growing structural databases make such derivations possible, although more rigorous strategies for building new efficient and widely applicable knowledge-based force fields remain one of the most challenging tasks of large scale molecular modeling in structural biology.

2.3.3. Structure-Based Models of Force Field. As pointed out in the Introduction, we know many experimentally

DOI: 10.1021/acs.chemrev.6b00163
Chem. Rev. 2016, 116, 7898−7936

determined protein structures. These could be used as a starting point for simulation studies. "Structure-based" models (SBMs), also called Gō-type models, employ a specific force field approximation.[174,175] Namely, only native-like interaction patterns, seen in a specific known and usually folded structure, are taken into account. In many cases, it is a significant simplification, since it is assumed that the folding process is directed by interactions that stabilize the known final structure. The folding intermediates of many proteins are certainly not necessarily native-like. Nevertheless, SBMs could be a quite useful tool for modeling near-native protein dynamics, especially when combined with non-native force field potentials.[176−180] There are also interesting modifications (extensions) of the SBMs patterns of interactions in which two, or another limited number, alternative basic structures are used for defining the subset of possible interactions, see section 4.2.1.

2.4. Sampling Schemes

Energy function transforms the flat world of conformational space accessible to a given biomolecular system into a very rugged hypersurface. The sampling scheme is a method of traveling through that hypersurface in search of desired conformations. Sampling schemes in biomolecular modeling were recently reviewed,[181−185] and here we only briefly describe the approaches commonly used in coarse-grained modeling. They belong to three broad categories: molecular dynamics (MD), Monte Carlo (MC), and heuristic approaches, such as Genetic Algorithm,[186] Taboo Search,[187] or Ant Colony Optimization.[188]

In MD, new configurations are generated by applying Newton's equations of motion to all atoms (or pseudoatoms) simultaneously over a small time step. This determines the new atomic positions and velocities and provides a trajectory describing how a given system evolves in time. In coarse-grained modeling, in which solvent is most often treated in an implicit fashion, collisions and friction forces should be introduced to mimic collisions of a solute molecule with its environment.[189] Since united coarse-grained atoms form a body typically larger than a solvent molecule, the Brownian motions theory is often applied to mimic the solvent effects. The stochastic effect of solvent molecules is introduced by a random displacement vector resulting from Brownian motions with zero mean and variance-covariance defined by a diffusion tensor. The frictional term, defined by Stokes' law, is often omitted, assuming that viscous force is much larger than the inertia tensor and the resulting formulation is referred as Brownian dynamics (BD). An important effect that must be included in BD algorithms to capture correctly the dynamics of the biomolecular chain is the hydrodynamic interaction (HI) between coarse-grained atoms. This in practice requires computing the square root of the $3N \times 3N$ diffusion tensor every time the tensor is updated during a simulation. This can be achieved with Cholesky decomposition as proposed by Ermak and McCammon in their pioneering work[190] on the BD algorithm with HI.

Discrete molecular dynamics (DMD), recently reviewed in refs 191 and 192, may be considered another particular way of solving Newtonian equations of motion. In this case, however, all energy wells are flat, and energy gradients (i.e., forces) are always equal to zero. DMD simulation assumes ballistic motion at constant velocities and searches for the closest collision event, saving computational time. In fact, the discrete, event-driven implementation was published before the "classical" time-resolved MD. DMD has been successfully applied for CG models to studying protein structure[193] dynamics[194] and aggregation.[195]

Monte Carlo (MC) methods essentially provide a random sample of conformations coming from the desired distribution. Boltzmann distribution, the most important case, is achieved when the Metropolis criterion is applied to accept or deny a new randomly created configuration. Typically, the new configuration is constructed as a small modification of the previous one. These may include translation and rotation of a randomly selected molecule in the system, or a small change to a subset of its internal degrees of freedom. Such a structural modification (also termed "MC move") must satisfy only very general rules: detailed balance and ergodicity. This opens up countless possibilities for introducing modifications that, while altering the structure as much as possible, attempt to avoid energy barriers and greatly reduce correlation between states (see Figure 4). Through a properly designed set of moves the Markovian property of the stochastic process leads to Markov chain Monte Carlo stochastic dynamics. Various aspects of the MC modeling of biomolecules have been recently reviewed.[196]

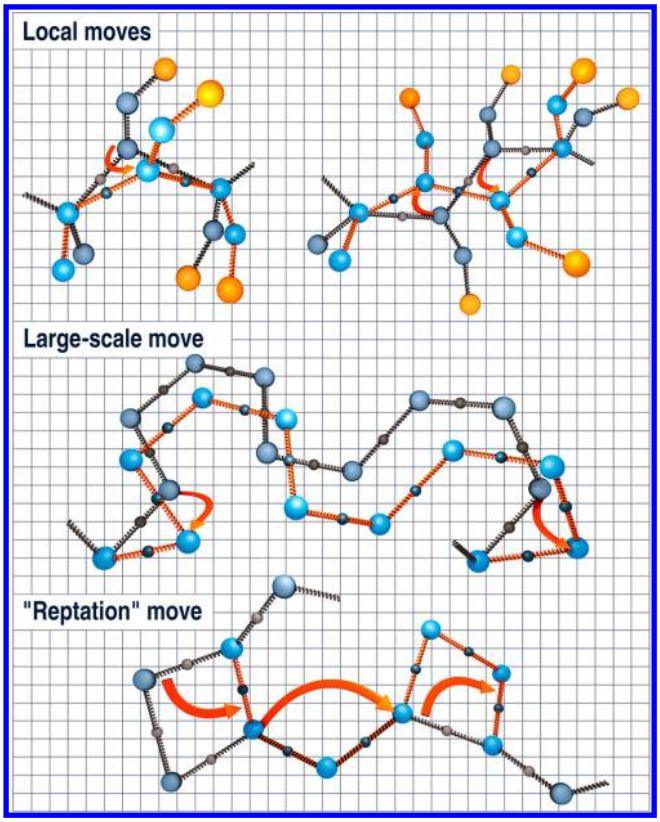


Figure 4. Example moves of a protein chain used for MC dynamics in the CABS model.[72] The upper panel shows local small-scale moves: two-bond and three-bond. The middle panel presents a large-scale move: small distance displacement of a chain fragment. The lower panel shows a "reptation" move in which a "bubble" on a protein chain is removed in one spot and randomly created somewhere else along the chain. A long random sequence of all moves provides an MC dynamics trajectory of the modeled protein chain. Large-scale and reptation moves are attempted less frequently than local moves. For clarity, the upper panel moves are shown with side chains (colored in orange), while the remaining panels without side chains. Positions of the alpha carbons are restricted to the underlying cubic lattice.

DOI: 10.1021/acs.chemrev.6b00163
Chem. Rev. 2016, 116, 7898−7936

Table 1. Selected Coarse-Grained Protein Models[a]

coarse-grained model	model design (representation of a single amino acid and force field design)	example applications, additional notes and availability
AWSEM (associated memory, water mediated, structure and energy model) developed by Wolynes, Papoian and co-workers[236]	Up to three-bead representation: Cα, Cβ, and O.	Model used for ab initio structure prediction of globular proteins,[236] prediction of dimerization interfaces of protein−protein complexes,[237] modeling the mechanisms of misfolding and aggregation[238,239] and the role of electrostatic effects in protein folding and binding.[240] AWSEM has been also extended to model α-helical transmembrane proteins.[241]
	Mixed (knowledge-based and physics-based) potential	Available as open source software at http://code.google.com/p/awsemmd/
Bereau and Deresno model, developed by Deresno and co-workers[242]	Up to four-bead representation: three backbone beads (N, Cα and C′) and one side chain bead located at Cβ.	Model used for studying protein folding[243] and protein aggregation.[242] Studies show that additional model tuning is needed to improve the stability of proteins with β-type or α/β-mixed secondary structure.
	Knowledge-based force field	Available as part of the ESPRESSO package:[244] http://espressomd.org/
CABS (C-alpha, c-beta, side chain) model, developed by Kolinski and co-workers[72]	Up to four-bead representation: Cα, Cβ, center of the side chain, and center of the peptide bond. Additionally, restriction of Cα positions to the cubic lattice (with 0.61 Å spacing) significantly speeds up the calculations.	Model used in template-based or ab initio protein structure prediction, validated in CASP competitions as one of the leading approaches;[171] loop structure prediction;[206] ab initio simulations of protein folding[101,207−209] and binding of intrinsically disordered proteins.[212] Used as a simulation engine in multiscale methods for protein structure prediction,[205] modeling of protein flexibility[210] and flexible protein-peptide docking.[213]
	Knowledge-based force field	Available with CABS-based tools at: http://biocomp.chem.uw.edu.pl/tools/
MARTINI model, developed by Marrink and co-workers[116]	Up to five-bead representation: one backbone bead (placed at the center of mass of the amino acid backbone), and up to four side chain beads.	Model originally developed for lipids[216,228] and subsequently extended to proteins.[116] Clearly the most popular model for the coarse-grained modeling of membrane proteins in the membrane environment. Its numerous successful applications are summarized in this review in section 4.5 and have been recently reviewed.[225]
	Physics-based force field	Available at: http://www.cgmartini.nl/, compatible with the GROMACS package:[245] http://www.gromacs.org/
OPEP (optimized potential for efficient protein structure prediction) model, developed by Derreumaux and co-workers[246,247]	Up to six-bead representation: full-atom for the backbone (N, HN, Cα, C′, and O) and single-bead for side chains (with an exception of proline having three beads).	Model used for protein folding;[247−250] aggregation studies;[251,252] structure prediction of peptides and small proteins;[253] modeling of the role of hydrodynamics in protein relaxation and peptide aggregation;[247] modeling of proteins, DNA-RNA complexes and amyloid fibril formation in a crowded environment;[246] *ab initio* peptide structure prediction.[254]
	Mixed (knowledge-based and physics-based) potential	Available with OPEP-based tools at: http://www-lbt.ibpc.fr/
PaLaCe (Pasi-Lavery-Ceres) model, developed by Lavery and co-workers[255]	Two-tier representation (one for bonded and another one for nonbonded interactions). Three backbone beads (N, Cα, and C′) are used for backbone representation and one or two beads for the side chain	Model used to maintain structures of folded proteins, and model their dynamic fluctuations and large-scale force-induced conformational changes;[255] protein flexibility prediction.[256]
	Physics-based force field	Available within the MMTK simulation package:[257]http://dirac.cnrs-orleans.fr/MMTK/
PRIMO model, developed by Feig and co-workers[106]	Up to seven-bead representation: three backbone beads (N, Cα, and combined CO) and one to five beads for the side chain (the representation was aimed to be sufficient for high resolution protein representation[105])	Model used in peptide and small protein structure prediction;[106] has been extended to membrane environments.[215]
	Physics-based force field	Available from the authors on request and distributed as part of the MMTSB Tool Set:[258] https://mmtsb.org/
Rosetta model, developed by Baker and co-workers[104]	Representation by all backbone atoms, Cβ and center of the side chain. Coarse-grained models are further refined in all atom representation (with explicit hydrogen atoms).	Model widely used for protein structure prediction, validated during CASP competitions as one of the leading approaches;[104] recent developments include improved protocols for high resolution refinement[46] and de novo blind predictions;[47] model implemented in numerous pipelines for protein−protein docking,[259] protein−ligand docking,[260] antibody modeling,[261] refinement of crystallographic structures,[262] refinement of NMR structures,[263] protein-peptide docking,[264,265] modeling of protein−DNA interactions.[266]
	Mixed (knowledge-based and physics-based) potential	Available at https://www.rosettacommons.org/ (RosettaCommons offers many web-interface servers for using Rosetta, including ROSIE, an easy-to-use web interface for selected Rosetta protocols[267]).
Scorpion (solvated coarse-grained protein interaction) model, developed by Basdevant and co-workers[268]	Up to three-bead representation: single backbone bead and one to two side chain beads.	Model initially developed for scoring protein−protein complexes.[111] Later, the protein model was combined with a water model and used for protein−protein recognition in a solvated environment of the barnase/barstar complex[268]
	Physics-based force field	
UNRES (united residue) model, developed by Liwo and co-workers[73]	Three-bead representation: Cα, peptide-group, and side chain.	Used in numerous protein folding studies;[269−272] protein structure prediction (successfully used in the CASP competition[273]); loop structure prediction;[274] protein−protein interactions;[275] protein−DNA interactions;[202] mechanisms of protein fibrillation,[276] large-scale rearrangements of protein complexes.[277]
	Physics-based force field	Available with accompanying tools at: http://www.unres.pl/

[a]Models are presented in alphabetical order.

Both MC and MD procedures typically have the same system setup including representation of molecules, definition of force fields, implementation of (periodic) boundary conditions, etc. However, it is quite difficult to provide direct comparison between the two. The equations used by MD remain always the same and possibilities to speed up computation lie mainly in handling constrains and numerical optimization. The Monte Carlo approach may use virtually any structural modification which does not violate its basic assumptions. MC moves crafted for studying a particular system can be devised. Nevertheless, various studies[197,198] indicate that MC is faster than MD, and sometimes many times faster. Another advantage of MC sampling is that the implementation of energy terms is not restricted to differentiable functions. This aspect may be important when the mathematical form of a force field is discrete, e.g. it is based on a histogram of a target property.

DOI: 10.1021/acs.chemrev.6b00163
Chem. Rev. 2016, 116, 7898−7936

Finally, in the case of computationally demanding high-resolution models, the sampling problem can be circumvented by running a large number of independent simulations and their further joint analysis.[199−201]

2.5. Examples of Protein Coarse-Grained Models

In Table 1 we present a brief overview of various coarse-grained protein models. For some of the models typical for certain classes of modeling strategies and/or very successful in some applications, we provide a wider description. The first two of the described models (UNRES and CABS) are quite universal, allowing studies of protein structure, dynamics and interactions. Many other coarse-grained models fit in this class. The third model, PRIMO, is particularly interesting because it assumes more accurate representation of protein structures and thereby, by somewhat higher simulation cost in comparison with the first two models, provides an almost atomistic picture of protein structure and dynamics. The other two, MARTINI and Rosetta, are probably the most widely used coarse-grained models. MARTINI is typically used as a simulation tool for membrane proteins, and Rosetta as a core module in a wide array of methods dedicated for different protein types, systems, and modeling tasks.

UNRES (united residues) is probably the most classical model of medium resolution, and it enables dependable and fast reconstruction of atom-level representation.[73] Three united atoms represent a single amino acid residue in this model (see Figure 2). The main chain is replaced by two atoms corresponding to Cα and the center of the Cα-Cα virtual bond and one pseudoatom of an ellipsoid shape of revolution representing specific side chains. Rotations of this ellipsoid mimic side chain conformational mobility. The force field of UNRES is rigorously derived from all-atom molecular mechanics models of interactions. Various MD and related methods have been used to sample the conformational space of model chains. The physics-based approach to force field and dynamic sampling enabled not only the structure prediction of small proteins but also the study of dynamics and interactions in larger systems.[202] Recently the range of UNRES applicability has been extended by careful implementation of structural restraints to some applications of the model.[202−204]

CABS (C-alpha, beta and side chain) is a medium resolution model. In comparison with UNRES, CABS provides similar resolution, but it is based on qualitatively different interaction and sampling concepts.[72] The choice of united atoms for modeling single amino acids is similar to that of UNRES except for the side chains which are represented by two spherical pseudoatoms, one centered on Cβ and the other placed in the center of mass of the remaining portion of the side chain, where applicable (see Figure 2). The main chain Cα positions are restricted to knots of a cubic lattice of small spacing, equal to 0.61 Å. This lattice Cα trace is used as the only independent variable that defines positions of other united atoms. The side chain positions are based on the local main chain Cα−Cα−Cα angle and the type of amino acid: the appropriate positions are derived from the statistical analysis of protein structures from the PDB database. Thanks to small fluctuations of the Cα−Cα distance and the rapid "pre-computing" of all possible local conformational transitions and associated changes of interactions, lattice representation enables extremely fast sampling of the conformational space, adding about 10-fold speedup of simulation compared to otherwise equivalent continuous space models. The force field of CABS is fully statistical, "knowledge-based" and probably quite unique, since interactions are treated as context dependent, and therefore take into account very complex multibody (including solvent) effects. For instance for single-domain globular proteins the interaction energy for contacts of two oppositely charged side chains depends on their mutual orientations. These interactions are strongly attractive for parallel orientations and weakly repulsive for antiparallel orientations. Indeed, polar amino acids are localized onto the surface of a globule, and thereby their close contact must be near-parallel. This regularity is reflected by statistical potentials derived from the analysis of a globular protein database. CABS sampling uses various Monte Carlo schemes. Since MC simulations use only local conformational changes (see Figure 4) long-time simulations mimic chain dynamics. A model has been successfully used in structure prediction (de novo and comparative modeling),[171,205,206] simulations of protein folding mechanisms[207−209] and flexibility of globular proteins,[210,211] and molecular docking.[212−214]

PRIMO (protein intermediate model) is of higher resolution than CABS or UNRES and provides another level of coarse-grained resolution, closer to atomistic representation.[105,106] The main chain is represented by three united atoms per residue, and the side chains by one or up to four united atoms, depending on their sizes and shapes. For instance, the phenylalanine side chain is represented by three pseudoatoms encoding size and orientation of the benzene fragment. This subtle level of coarse-graining is sufficient for ensuring noticeable simulation speedup in comparison to all-atom simulations. The force field of PRIMO is very rigorously constructed and maintains the structure of standard MD interaction models. The scaling of the force field has been done in a way that provides good transferability of protein systems. What is important, the PRIMO model enables productive studies of the solvent effect on protein dynamics, including simulations of membrane proteins[215] and reproducible folding simulations of small proteins.

The MARTINI coarse-grained model was initially designed by the Marrink group just for membranes composed of lipid molecules.[216] Further versions were subsequently extended to include peptides, proteins[116,217] and other small biomolecules.[217,218] Schulten[219] and Sansom[220] groups have also developed other protein coarse-grained force fields, compatible with the MARTINI lipid force field. Others also proposed several modifications to the model.[221−224] MARTINI modeling tools are now continuously refined by the research groups of Marrink and Tieleman.[116,225] The MARTINI force field is based on one-to-four mapping, which means that on average four heavy atoms including associated hydrogens are represented by a single coarse-grained bead (see Figure 5). Consequently, one coarse-grained water bead corresponds to four water molecules. One coarse-grained ion bead mimics a single ion including its first hydration shell. Small ring-like fragments (e.g., aromatic amino acid side chains) or small molecules (e.g., benzene, cholesterol) are mapped with slightly higher resolution of up to two heavy atoms per one coarse-grained bead. To properly reproduce the chemical nature of the modeled systems, four main types of coarse-grained particles are defined: polar (P), nonpolar (N), apolar (C), and charged (Q). The four main types of coarse-grained particles are divided into subtypes based on hydrogen-bonding capabilities (donor, acceptor, both or none) and polarity (ranging from 1 = low polarity to 5 = high polarity) giving a total of 18 unique "building blocks". The described mapping scheme provides a

DOI: 10.1021/acs.chemrev.6b00163
Chem. Rev. 2016, 116, 7898−7936

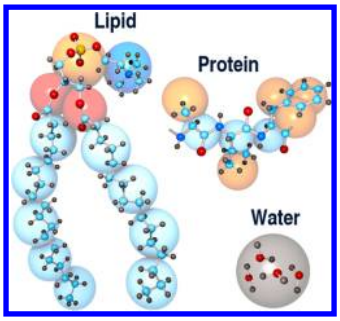


Figure 5. All-atom versus coarse-grained representation in the MARTINI model.[225] All-atom representation is shown in balls and sticks, while coarse-grained representation in large spheres. The figure shows an example lipid molecule, fragment of a protein chain, and water representations.

relatively straightforward and effective way of switching from all-atom to coarse-grained representation for a wide range of biological systems. Interactions between coarse-grained particles are described by a force field containing terms typical for other classical force fields. Nonbonded interactions are controlled by a Lennard-Jones (LJ) 12-6 potential, where ε_{ij} depends on types of interacting coarse-graining particles. Electrostatic interactions are defined by the Coulombic energy function. The nonbonded parameters have been adjusted to reproduce experimental thermodynamics data of the free energy of hydration, free energy of vaporization and partition free energies between water and a number of organic phases for each of the 18 types of coarse-grained particles.[225] Interactions describing bond lengths, angles and dihedrals are controlled by a standard set of potential energy functions. Parameters were calculated based on structural data derived either from atomistic geometry or from an iterative procedure in which parameter values for coarse-grained representation were systematically adjusted to obtain satisfactory overlap with the distribution function resulting from all-atom MD simulations of corresponding atom groups. The practical application of the method is facilitated by a simple mapping procedure based on the "building block" concept, which allows generation of a unified set of parameters and topologies for systems of different types. What is very useful, MARTINI provides parameters for a large number of molecules including different lipid types, sterols, sugars, peptides, polymers, and more. The force field was originally developed to be used in the GROMACS simulation package[226] but the general form of the potential energy function also allowed its implementation in other well-known MD simulation codes such as Desmond,[227] GROMOS,[228] and NAMD.[219]

The Rosetta[104,229] model uses two protein representations: coarse-grained (see Figure 2) and all-atom, with all hydrogen atoms present. Rosetta defines protein conformation in the dihedral space. Thus, a coarse-grained polypeptide chain has three degrees of freedom (phi, psi and omega) for each amino acid residue. All-atom representation also includes Chi angles for side chains. All the other internal degrees of freedom are fixed to "ideal" values, although they might be relaxed in some applications, e.g., at the final stage of high-resolution structure refinement. Unlike other approaches to protein modeling, the two representations in Rosetta are tightly connected and the program seamlessly switches from low to high resolution. The coarse-grained model includes all heavy atoms of the backbone, beta-carbons and virtual atoms representing amino acid side chains. A few dozens of energy terms have been defined in Rosetta depending on resolution (either coarse-grained or all-atom), experimental data included in modeling and the specific problem under study. The Rosetta source code is organized in a hierarchical manner.[229] Low-level classes provide procedures typical for macromolecular modeling such as evaluation of scoring terms or altering internal degrees of freedom. They are combined into protocols. Ab initio,[230] the historically first Rosetta protocol, is used to predict protein structure based solely on its sequence information. Modeling is conducted in coarse-grained representation. Conformational space is limited also by the sampling scheme. To create a new conformation from the previous one, a randomly selected fragment of known protein structure substitutes a local structural fragment. The fragments themselves are extracted from a nonredundant set of proteins based on sequence and secondary structure similarity. Three and nine amino acid fragments were originally used, but in the current implementation[231] their length is not restricted. After every MC move, Cartesian coordinates must be recovered for energy evaluation. The fragment assembly simulation is conducted in a hierarchical manner. In the first stage only nine residue fragments are used for sampling and the energy function is limited to VdW, hydrogen bonding and collapsing terms. Further stages introduce smaller changes to the structure while the energy function becomes more elaborated. Unlike other methods used for protein modeling, a single modeling Rosetta run results in a single low-energy conformation. The protocol must be therefore repeated many times to gather proper statistics. This approach, however, ensures that the conformations are statistically uncorrelated. Coarse-grained conformations are further subjected to side chain reconstruction and all-atom energy minimization (FastRelax protocol). To sample the conformational space, Rosetta uses several approaches to alter dihedral DOFs: fragment insertions, backrub moves,[232,233] rotamer library[234] or small perturbations of particular internal coordinates. For example, the Rosetta ProteinDesign protocol[235] is based on the Monte Carlo search strategy to optimize a protein sequence by mutating one residue at a time. During a single step of the ProteinDesign protocol, backbone DOFs remain fixed and side chain Chi angles are modeled and assigned according to the rotamer library. This step is followed by backbone relaxation to accommodate the designed amino acids. A number of Rosetta applications are also outlined in Tables 1 and 3.

3. MULTISCALE MODELING

3.1. Example Strategies

In principle, every simulation that allows transfer of information between at least two different levels of granularity can be considered multiscale. Multiscale methods are more efficient, and enable analysis of larger systems in a longer time scale with a simultaneous ability to preserve a high level of details when necessary. This idea was applied to biological objects perhaps for the first time by Levitt and Warshel in 1976 in their study of mechanisms of enzyme action.[35] Since then multiscale modeling has found successful applications in the modeling of proteins,[25,38,39,278] membranes,[279,280] ribonucleic acids,[281−283] and large protein−protein or protein−membrane complexes.[38,284−287] Depending on the given scientific problem, several combinations of methods have been proposed. In practice, the most common are QM/MM (quantum mechanics/molecular mechanics) and all-atom/coarse-grained.

DOI: 10.1021/acs.chemrev.6b00163
Chem. Rev. 2016, 116, 7898−7936

QM/MM modeling was historically the first multiscale approach used for chemical computation. Currently there are many different versions of the QM/MM approach; however, all of them combine QM calculations for the active site which is submerged in a simplified environment (full protein or solvent).[25] With time, QM/MM has become the most important computational method for studying enzyme function or other biomolecular processes that involve changes in electronic structure,[288−293] and it is the basis of modern enzymology.[294] For more information concerning the use of QM/MM, please refer to relevant reviews.[25,38,39,295−298]

The all-atom/coarse-grained multiscale modeling approach (see Figure 6) has emerged as one of the most promising tools

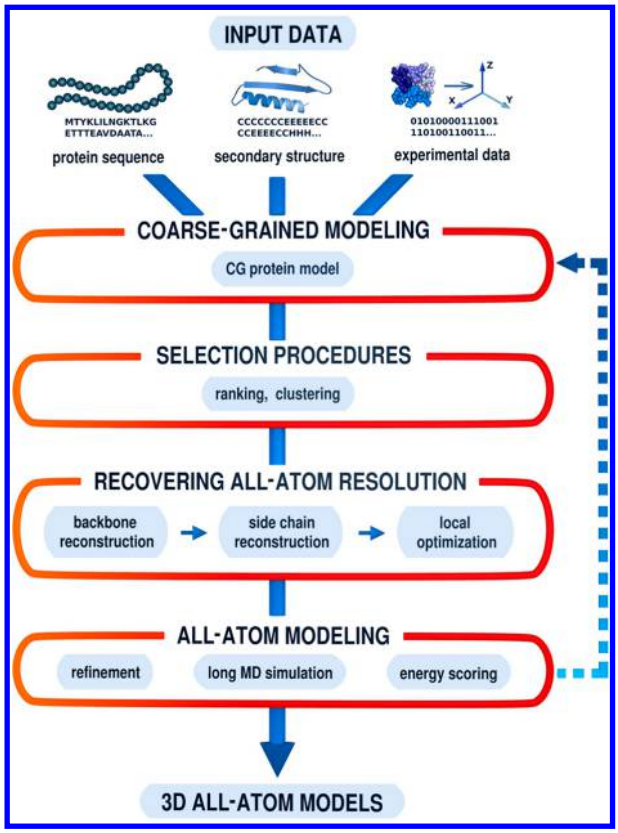


Figure 6. Typical multiscale modeling scheme that merges coarse-grained and all-atom modeling. In specific tasks, the resulting all-atom structures could be used as an input for the next stage of coarse-grained simulations. Other multiscale schemes are briefly discussed in the text.

of computational biology, and it combines the efficiency of coarse-grained simulations and details of all-atom simulations for the characterization of a broad range of molecular systems.[299] In this methodology all-atom and coarse-grained energy is frequently calculated as the sum of atomistic, coarse-grained, and hybrid parts of the system:

$$E = E_{AA} + E_{CG} + E_{AA/CG} \quad (5)$$

where the terms for every part may be evaluated in different ways.[300]

The coarse-grained potential is often highly simplified, and even electrostatic forces are mostly neglected.[301] The problem of all-atom/coarse-grained modeling has been addressed quite some time ago by Skolnick and co-workers.[302] In their study on folding pathways of the leucin zipper they discovered that the accuracy of coarse-grained modeling could be significantly improved when final structures from low-resolution simulations were additionally refined with a detailed atomistic model. This observation and further studies[303] led to a hypothesis that it should be achievable and beneficial to develop hybrid methods for protein structure prediction. Another study undertaken by Warshel and co-workers[304] also used a simplified potential as a reference potential for calculating all-atom free energies.

We can distinguish multiscale methods in which coarse-grained simulations are used at the initial stages of the modeling process to provide data for further all-atom simulations[44,299,304,305] (see Figure 6), or conversely, information from all-atom simulations is transferred to coarse-grained models.[125,306,307] Such a strategy is called "serial multiscale modeling".[37] There is also another group of methods which may be called "parallel multiscale modeling" that combine fine-grained and coarse-grained representations in a single mixed-resolution simulation.[308−310] A typical example of parallel multiscale modeling may be simultaneous calculation in both resolutions: simultaneous use of a coarse-grained model to identify the area of possible conformations with an all-atom method to improve accuracy of the resulting model.[311] A widely used strategy of parallel multiscale modeling is also treating a large part of the system with a coarse-grained model while the other, smaller section is treated with atomic resolution.[25] Some interesting applications of serial and parallel multiscale modeling are provided in section 4.1.

Successful multiscale modeling, regardless of the type, needs efficient and reliable algorithms for transferring information between calculations with different resolutions.[25] Multiscale dynamic modeling is even more demanding since proper calculation ought to be fast enough so as not to hinder the benefit of coarse-graining. To date many different approaches that define the concept of a boundary have emerged. However, since information exchange methods are the key limiting factors, new theories that would fill in these gaps are still of great need. The boundary between different levels of modeling can be arranged with fixed resolution as shown in the studies of lipid bilayer permeability to small molecules,[312] the large-scale motion impact on the function of outer membrane protease T[313] or membrane-bound ion channel studies.[308] Another approach is to make them adaptive,[314] which means that it is allowed to change granularity of a selected part of the system during the simulations on demand. Different methods for flexible linking have been developed over time: AdResS,[315] Adaptive partitioning,[316] Hot spot method[317,318] and ONIOM-XS.[319] However, despite the efforts, there is still a need for developing better and faster algorithms that would allow changing the resolution on-the-flow.

3.2. Reconstruction of Atomic Representation from Coarse-Grained Models

Information transfer from coarse-grained to atomic representation is tightly connected with the task of reconstruction of atomic details (see Figure 6). The problem of determining protein structure only from the Cα trace or coarse-grained representation is not new and can be found in literature multiple times, and thus most algorithms are based on solutions developed originally for protein structure prediction, homology modeling or protein design.[299]

The process of protein reconstruction to atomistic structure can be divided into two separate steps: rebuilding the backbone and adjusting the side chain atoms. Several different approaches

DOI: 10.1021/acs.chemrev.6b00163
Chem. Rev. 2016, 116, 7898−7936

Table 2. Example Reconstruction Algorithms

program name	realized reconstruction task	additional comments and availability
BBQ (backbone building from quadrilaterals)[322]	BBQ performs main chain reconstruction from the C-alpha trace.	Method designed for robust and efficient backbone reconstruction. Tested on 81 nonredundant protein sets derived from PDB and generated near-native coarse-grained decoys.
	Uses a library of backbone quadrilaterals.	Available as a standalone program, part of the Bioshell package.[349,350]
Modeller[351,352]	Modeller enables reconstruction and refinement from the C-alpha trace or models with missing atoms.	The reconstruction procedure is available as part of the popular Modeller package for comparative modeling.
	Reconstructs an all-atom model based on a protein template in coarse-grained representation. Energy minimization and all-atom scoring may follow the reconstruction procedure.	
ModRefiner[102]	ModRefiner performs all-atom reconstruction from the C-alpha trace.	Method designed to handle unphysical local distortions in coarse-grained models and to improve the physical quality of a local structure.
	After backbone reconstruction, ModRefiner optimizes the backbone H-bond network. Side chain reconstruction is followed by additional optimization with a composite physics- and knowledge- based force field.	Has an option to drive the refinement simulation toward the desired secondary structure.
		Available as a server or a standalone program. The standalone program has an option of *ab initio* structure refinement.
NCN[353]	NCN performs side chain reconstruction from the protein backbone.	Highly accurate algorithm for side chain reconstruction. Tested on 65 high resolution X-ray structures. Available as a standalone program.
	Uses optimized OPLS parameters, simulated-annealing search strategy and a detailed rotamer library.	
OPUS_Rota[354]	Opus_Rota performs side chain reconstruction from the protein backbone.	Accurate method for side chain reconstruction. Tested on 65 high resolution X-ray structures and the Wallner and Elofsson homology-modeling benchmark set.[355] Available as a stand-alone program.
	It combines commonly used potential terms, solvation energy with unique orientation-sensitive potential (OPUS-PSP) and the Monte Carlo sampling scheme.	
OSCAR[356]	OSCAR methods (-o, -star) perform side chain reconstruction from the protein backbone.	Accurate methods designed for side chain reconstruction to obtain a realistic all-atom protein model. Tested on 218 proteins and a RAPPER decoy set (loop side chains reconstruction). Available as a standalone program.
	Combine accurate, orientation-depended, optimized side chain atomic energy with a flexible (OSCAR-o) or rigid (OSCAR-star) rotamer model.	
PD2[357]	PD2 performs main chain reconstruction from the C-alpha trace.	Method tested in all-atom reconstruction tasks in combination with Rosetta and SCWRL4.0.
	Uses a library of short fixed length backbone fragments for constructing the structural alphabet with a Gaussian mixture model.	Available as a server or a standalone program. The program features an optional energy minimization step.
Pulchra[323]	Pulchra performs all-atom reconstruction from the C-alpha trace.	Method designed for fast and robust calculations. Pulchra accepts even seriously distorted input structures. Tested on 500 random decoy structures from the prediction benchmark.
	Uses a simple force field and steepest-descent minimization for backbone reconstruction based on a modified algorithm described by Milik et al.[325]	Available as a standalone command line application.
RACOGS[299]	RACOGS performs all-atom reconstruction from the C-alpha trace.	Method designed for multiscale calculations and optimized to obtain a physically realistic all-atom model from coarse-grained models. Tested on 606,000 coarse-grained structures of the wild src-SH3 domain and ribosomal protein S6 and a misfolded mutant as well as on a subset of 2945 PDB structures. Available as a web server.
	Uses backbone reconstruction based on the work of Feig[324] and Milik[325] with an efficient side chain reconstruction method by Xiang and Honig.[358] Further addition of hydrogen atoms and all-atom minimization is performed with AMBER 8.[359]	
REMO[360]	REMO performs all-atom reconstruction from the C-alpha trace.	Method designed for refining I-TASSER coarse-grained models. Tested on 230 nonredundant proteins as well as coarse-grained models. Tested in a blind test in CASP8. Available as a web server and a standalone application.
	Removes steric clashes in the C-alpha trace and rebuilds backbone heavy atoms from a backbone isomer library. The method predicts the hydrogen bond network and uses an SCWRL algorithm for side chain reconstruction.	
Rosetta (catoalla-tom program)	Performs all-atom reconstruction from the C-alpha trace.	Program available as part of the Rosetta modeling suite.[104] Conformational search is guided by Rosetta energy functions and, optionally, by experimental data such as EM maps.
	Uses a library of fragments extracted from known structure fragments by combining them as rigid bodies in the Cartesian space.	
SABBAC[361]	SABBAC performs backbone reconstruction from the C-alpha trace.	Method tested on a subset of proteins from PDB. This robust method provides an answer even for degenerated C-alpha traces. Available as a web server.
	Uses a small library of short fragments extracted from known protein structures and a greedy algorithm for fragment assembly and optimization.	
SCATD[362]	SCATD performs side chain reconstruction from a protein backbone.	Method tested on 180 proteins (the same benchmark set as for SCWRL 3.0). Available as a standalone application.
	Uses the rotamer library from SCWRL (version 3.0). SCATD computes interaction scores between atoms, uses dead-end elimination criteria and energy minimization via tree decomposition.	
SCWRL (version 4.0)[363]	SCWRL performs side chain reconstruction from a protein backbone.	Method widely used in protein modeling pipelines. Tested on 379 proteins from PDB.
	Uses a backbone dependent rotamer library, calculates self-and pairwise energies, builds and solves graphs with a modified tree decomposition algorithm.	Available as a standalone program and a dynamic-linked library for incorporation into other software programs.
SidePRO[343]	SidePRO performs side chain reconstruction from the protein backbone.	Method designed for fast side chain reconstruction. Tested on 379 proteins (SCRWL4 benchmark set), 94 proteins from the CASP9 data set and 7 protein complexes.

DOI: 10.1021/acs.chemrev.6b00163
Chem. Rev. 2016, 116, 7898−7936

Table 2. continued

program name	realized reconstruction task	additional comments and availability
	Uses a knowledge-based energy function and artificial neural networks.	Available as a web server and computer code.

have been proposed to solve the first task. These algorithms may use analytical methods,[320,321] statistical propensities[322−325] or even whole short peptide fragments[326,327] extracted from known structures. Reliable side chain reconstruction is a more difficult task. It has been shown that the side chain positioning problem is NP-complete,[328] which means that the complexity of the calculation rises exponentially and no exact polynomial-time algorithms are known.[329] Most of the methods for side chain assignment use a rotamer library[330,331] built from known structures and an energy function that allows finding a global energy minimum. Many methods for side chain reconstruction have been proposed: Monte Carlo sampling,[332,333] dead-end elimination,[334−336] simulated annealing,[337] local optimization,[330] genetic algorithms,[338,339] integer linear programming,[340] graph decomposition,[341,342] and other combined approaches.[329,343−348] In Table 2, the currently available methods are briefly outlined. Finally, it needs to be noted that protein conformations generated by coarse-grained models may exhibit some small unphysical distortions that are typical for specific coarse-grained models. Those distortions may be not well tolerated by a reconstruction algorithm. Therefore, before the application, the chosen reconstruction algorithm should be tested in combination with a particular coarse-grained model (see also the related comments in column 3 of Table 2).

4. APPLICATIONS

4.1. Coarse-Grained Models in Multiscale Modeling Pipelines

Over the years many research groups have undertaken studies on the use of the all-atom/coarse-grained multiscale modeling approach trying to answer different biophysical questions and develop highly specific tools. As described in section 3.1, these tools can be divided into parallel and serial multiscale modeling approaches. The parallel approaches usually require additional strategies for the integration and exchange between the different levels of resolution, while serial multiscale approaches are more common and straightforward.

MMTSB[258] is a good example of a toolset that enables custom parallel multiscale simulations. It combines packages that allow simulation with all-atom resolution such as CHARMM[113] or AMBER[364] combined with the coarse-grained modeling approach, MONNSTER[365] (based on a medium resolution SICHO lattice model[95]). This approach was successfully tested for scoring protein conformation, peptide folding and prediction of missing protein fragments.[258] Another computationally complex task for which the parallel multiscale approach has been successfully applied is modeling interactions of proteins submerged in solvents or lipid membranes. In this case solvents or membranes can be represented in a coarse-grained manner while proteins are treated with atomistic resolution.[20] In one of such force fields, named PACE (protein in atomistic details coupled with coarse-grained environment) and developed by Han et al.,[366] the united atom representation of a protein is combined with the MARTINI coarse-grained model of a solvent or membrane. In the folding simulations the authors obtained satisfactory coherence with experimental data, comparable to fully atomistic approaches.[367] An interesting parallel multiscale approach has been also described by Machado et al. for simulating nucleic acids.[300] In this model the region of interest is treated with all atom details using the AMBER force field[368] while the rest of the system is approximated by coarse-grained representation.[369] For interaction calculations the method employs a common Hamiltonian function that serves for both the all-atom and the coarse-grained simulation level and makes further extension to QM/MM or a different level of coarse-granularity uncomplicated. Another parallel model that allows using the multiscale approach to simulate the crowding effect on peptides is described by Predeus et al.[370] In this simulation crowders (proteins G) are represented with the PRIMO (protein intermediate model)[105] coarse-grained model, while peptides of interest are accounted for in atomistic detail using the CHARMM force field.[113,114]

Among serial multiscale methods it is worth to mention a model developed by Zacharias principally for studying protein−protein interactions.[371] It combines all-atom GROMOS[372] and coarse-grained ATTRACT[373] force fields. It allows for the exhaustive Monte Carlo sampling of coarse-grained representation of the interacting proteins followed by an accurate all-atom description of favorable protein−protein complex geometries. Another method that combines the GROMOS force field with the coarse-grained OPEP potential[374] was presented in the study of the amyloid-forming peptide.[252] In this example of serial multiscaling, coarse-grained simulation was used to simulate aggregation of amyloid peptides, and then the stability of the derived species was tested with all-atom resolution.

Finally, among multiscale all-atom/coarse-grained tools there is a growing number of methods available as web servers that are especially useful for nonexpert users. This trend of making molecular modeling methods available as easy to use and accessible web servers (so-called "serverification"[267]) is expected to continue. Below we provide examples of multiscale modeling servers that use coarse-grained modeling tools. More detailed information concerning all-atom/coarse-grained multiscale modeling is available in recent reviews.[39,375,376]

4.2. Simulations of Protein Dynamics

So far, structural biology has focused mainly on the analysis of static X-ray structures. Recent advances in experimental techniques strongly indicated that protein function is in many cases dictated by dynamics.[5,6] Thus, the classical "structure determines function" dogma has been extended to include dynamics. Consequently, the ultimate goal of contemporary structural biology is to add time, the fourth dimension, to the characterization of protein structure. Unfortunately, a protein energy landscape is highly multidimensional and tied to a specific set of conditions (e.g., temperature, pressure, and environment specifics). Therefore, the characterization of protein dynamics, either by theory or by experiment, is very challenging and the difficulty of the problem increases with the scale of mobility and size of protein systems.

4.2.1. Protein Folding and Large Scale Dynamics. For most proteins, the time scale of protein folding is beyond the possibilities of all-atom MD with explicit solvent.[16,17] Recent

DOI: 10.1021/acs.chemrev.6b00163
Chem. Rev. 2016, 116, 7898−7936

Table 3. Examples of Multiscale Modeling Web Servers

server name (reference)	coarse-grained modeling task	method summary, availability
		Protein Structure Prediction
I-TASSER[48,173,377]	Uses a coarse-grained protein model for the *ab initio* modeling of unaligned regions (mainly loops).[170]	I-TASSER performs comparative and *ab initio* prediction of protein structure. The method has been validated during several recent CASP competitions as the leading approach for protein structure prediction.[48,173,377] Available at: http://zhanglab.ccmb.med.umich.edu/I-TASSER/
CABS-fold[205]	Uses a CABS coarse-grained model as a conformational search engine.	CABS-fold performs *ab initio* and consensus-based prediction of protein structure. The methodology was validated during the CASP competition as one of the leading approaches. Available at: http://biocomp.chem.uw.edu.pl/CABSfold/
Robetta[378]	Uses coarse-grained Rosetta representation in the initial stage of structure prediction.	Robetta provides *ab initio* and comparative modeling prediction of protein structure. The method has been validated during several recent CASP competitions as one of the leading approaches. Other capabilities include prediction of the effects of mutations on protein−protein interaction, protein design and protein−protein docking. The server can also utilize NMR constraints data. Available at: http://robetta.bakerlab.org/
Phyre2[379]	Uses a Poing[380] coarse-grained model for *ab initio* modeling.	Phyre2 performs comparative and *ab initio* prediction of protein structure. The method has been validated as one of the leading approaches in CASP competitions. The server provides a suite of tools to analyze protein structure, function and mutations. Available at: http://www.sbg.bio.ic.ac.uk/phyre2
Pep-fold[254]	Uses an OPEP coarse-grained model as a conformational search engine.	Pep-fold performs *ab initio* structure prediction of peptide structure (between 9 and 36 amino acids). The server allows user specified constraints such as disulfide bonds and inter-residue proximities. Available at: http://bioserv.rpbs.univ-paris-diderot.fr/services/PEP-FOLD/
		Protein-Peptide Docking
CABS-dock[213,214]	Uses a CABS coarse-grained model as a conformational search engine.	CABS-dock performs flexible protein-peptide docking without prior knowledge about the binding site. During on-the-flydocking, CABS-dock allows full flexibility of the peptide and moderate flexibility of the protein receptor structure. The method has been tested on a large benchmark set of protein-peptide complexes and has proven to be effective in building high and medium accuracy models. Available at: http://biocomp.chem.uw.edu.pl/CABSdock/
pepATTRACT[381]	Uses an ATTRACT coarse-grained model for rigid body docking and scoring.	pepATTRACT performs flexible protein-peptide docking without prior knowledge about the binding site. The best scored complexes are subjected to atomistic refinement using the iATTRACT[382] procedure and all-atom MD. The protocol was tested on 80 peptide−protein complexes.[381] The web interface enables generation of input docking scripts and the docking can be performed on the user's machine with a local installation of the ATTRACT program. Available at: http://www.attract.ph.tum.de/peptide.html
Rosetta FlexPepDock[264,383]	Uses coarse-grained Rosetta representation in the initial stage of docking.	Rosetta FlexPepDock is a high-resolution docking and refinement protocol for modeling protein-peptide interactions. The server requires an approximate specification of the peptide binding site (anchor residue). The method has been tested on a large benchmark set of protein-peptide complexes and has been shown to generate high-resolution models.[265] Available at: http://flexpepdock.furmanlab.cs.huji.ac.il/index.php
		Modeling the Flexibility of Globular Proteins
CABS-flex[210]	Uses a CABS coarse-grained model as a fast simulation engine.	The method has been shown to generate a similar picture of protein flexibility compared to all-atom MD[211] and NMR ensembles.[384] Available at: http://biocomp.chem.uw.edu.pl/CABSflex/
NMSim[385]	Uses coarse-grained normal-mode analysis	The method allows performing three simulation types: unbiased search of the conformational space; pathway generation by targeted simulation; and radius of gyration-guided simulation. NMSim has been shown to be a computationally efficient alternative to all-atom MD. Available at: http://www.nmsim.de
FlexServ[386]	Uses three coarse-grained algorithms for simulations of protein flexibility: discrete dynamics, normal-mode analysis and Brownian dynamics.	FlexServ allows complete analysis of flexibility using a large variety of metrics. The server can also analyze user provided trajectories. Available at: http://mmb.pcb.ub.es/FlexServ/
		Protein Design and Interactions
ATTRACT[387]	Uses the ATTRACT coarse-grained model for protein−protein docking	The web service supports systematic rigid-body protein−protein docking, as well as various kinds of protein flexibility.[387] ATTRACT has been successfully used to predict complex structures in various rounds of CAPRI.[388−390] The web interface enables generation of input docking scripts and the docking can be performed on the user's machine with a local installation of the ATTRACT program. Available at: http://www.attract.ph.tum.de/
BeAtMuSiC[391]	Uses statistical potentials[392,393] adapted to coarse-grained protein representation to evaluate the energetic change induced by the mutation.	BeAtMuSiC allows fast assessment of changes in binding affinity between two proteins in the complex caused by point mutations. The method was evaluated within the 26. CAPRI round and stood among the top performers for the analysis of ∼2000 mutations in two designed inhibitors of influenza hemagglutinin.[391] Available at: http://babylone.ulb.ac.be/beatmusic/
DOCK/PIERR[394]	Uses coarse-grained scoring functions for protein docking.[395]	DOCK/PIERR performs docking of proteins given their individual tertiary structures. The docking algorithm has been tested on docking benchmark data sets and has proven to perform similarly as other state-of-the-art methods.[396] The server has ranked fourth in the 2013 CAPRI server assessment. Available at: http://clsb.ices.utexas.edu/web/dock.html
ENCoM[397]	Uses coarse-grained normal-mode analysis.	ENCom allows prediction of the effect of point mutations on function, thermostability and dynamics of proteins with multiple chains. In addition, the method generates comprehensive, geometrically realistic conformational ensembles of mutated proteins. Available at: http://bcb.med.usherbrooke.ca/encom.php

DOI: 10.1021/acs.chemrev.6b00163
Chem. Rev. 2016, 116, 7898−7936

Table 3. continued

server name (reference)	coarse-grained modeling task	method summary, availability
		Protein Design and Interactions
RosettaDock[398]	Uses coarse-grained Rosetta representation in the initial stage of docking.	RosettaDock predicts the structure of protein−protein complexes, such as antigen−antibody pairs, enzyme−inhibitor pairs or regulatory proteins. It has been successfully validated in the CAPRI blind challenge on diverse targets, also in combination with the Funhunt classifier[399] for selection of low-energy conformations close to the native conformation from other low-energy ensembles. Available at:http://graylab.jhu.edu/docking/rosetta/
RosettaAntibody[261,400]	Uses coarse-grained Rosetta representation in the initial stage of modeling.	RosettaAntibody uses a comparative modeling approach to build structures of complementarity determining regions (CDRs). In the second stage, the CDR and H3 loop is remodeled *ab initio* and orientation of VL/VH domains is optimized using a Rosetta protocol. Paratope side chains and loop backbones are refined simultaneously. The procedure has been tested during the Antibody Modeling Assessment II experiment.[261] Available at: http://antibody.graylab.jhu.edu/antibody

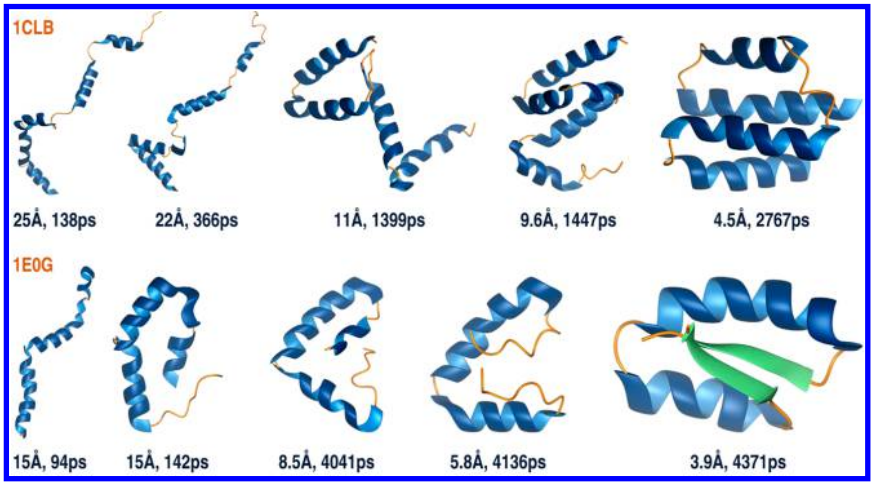


Figure 7. Folding pathways obtained in ab initio simulations using the UNRES coarse-grained model and Langevin dynamics.[269] The pathways are presented for two proteins 1CLB and 1E0G. The protein models are marked with the RMSD values (root-mean-square deviation from the corresponding experimental structures, in Ångstroms) and simulation time. The simulations took a few hours of a single CPU time; therefore, the UNRES model provided a three to four-order-of-magnitude speed-up relative to all-atom MD simulations.

advances in building specialized hardware dedicated to MD simulations enable us to reach the 1 ms time scale for a small protein,[401] which means significant (100 fold) speedup in comparison to the previous 10 μs record.[402] Because of the computational cost, unbiased (using no knowledge about the native structure) MD simulations of protein folding with explicit solvent are limited to small (up to ∼100 amino acids) fast-folding proteins,[16,401,402] while simulations of larger proteins are restricted to studies of near-native dynamics or high temperature unfolding. A few protein folding pathways have been described in detail at atomic resolution, thanks to combining experimental data with all-atom MD techniques.[403,404] Coarse-grained models offer significant extension of the simulation time scale. For example, compared to all-atom MD with explicit solvent, the speedup can be between 10^3 and 10^4 times for the UNRES model[271] or the CABS model[211] and even 10^7 times for much simpler models.[27]

In the past two decades, the field of protein dynamics simulation was dominated by the native-centric view of protein folding.[200] This view was supported by experimental studies showing that, on a general level, folding can be described as a simple two-state process, and that transition state structures are very native-like. The assumption that native interactions only are sequence-dependent ensures that the native structure is always in the energy minimum. Introducing this assumption into molecular models offered the possibility of significant simulation speedup and unification of the protein folding picture.[174,175,177,178] The models with such an interaction scheme, called structure-based models (SBM) or Gō models, have become widely used in simulations of folding mechanisms, folding kinetics and functional motions.[177−180,405−407] Many variations of SBMs proved to be valuable in descriptions of conformational dynamics, which is interpreted by the native-like character of the key transition states. However, the SBM-based approach seems to be increasingly unsatisfying in the context of emerging studies on a significant role of non-native interactions in protein folding.[408−410] Importantly, functions of many proteins are associated with the existence of multiple different conformations. Thus, it is obvious that building an SBM force field based on a single native state structure may be insufficient for the complete exploration of a conformational landscape.[411] Therefore, modern SBMs have been extended by adding information about a few different conformational states (multibasin models).[411−413] Multibasin SBMs are built on the same concept as single SBMs with the exception that they use more than one conformational state (for example, two distinct structures crystallized in different, bound and unbound, states may be used) as references for simplified interaction patterns. SBMs are also extended by using additional information, for example enforcing protein motions[414] or coevolutionary information extracted from multiple sequence alignments.[411] Regardless of the extension of SBMs, the fundamental question arises of whether the interaction model based on a specific structure, or selected structures, is sufficient in a particular case to describe the real functional dynamics.

In principle, deep understanding of protein folding mechanisms requires simulation models that are not biased by structural information about the final native state, and that

DOI: 10.1021/acs.chemrev.6b00163
Chem. Rev. 2016, 116, 7898−7936

do not exclude different intermediates than fragments of the native-like structure. Based on the earlier advances (see reviews in refs 19−21 and 44), in recent years we have witnessed protein (or peptide) folding mechanism studies using various coarse-grained models that go beyond the SBM interaction scheme: UNRES[269−272] (see Figure 7), OPEP,[247−250] PACE,[415,416] the model by the Voth group,[417,418] the Bereau and Deserno model,[243] CABS,[101,207−209] or the TerItFix approach.[419,420] These examples demonstrate the usefulness of models based on physics-based potentials, physics-based potentials combined with terms from statistical analysis of folded structures or solely on statistical potentials (CABS, TerItFix). While physics-based models present a straightforward approach to the investigation of folding dynamics, the application of statistical potentials raises questions about the validity of such an interaction scheme to the simulation of denatured structures. Nevertheless, it has been demonstrated that the evolution of folding events, from denatured to near-native states, guided by statistical potentials is consistent with experimental data[207,208,419,421] or with all-atom MD simulations for small fast-folding proteins.[420] Consequently, these results strongly suggest that the nature of interactions (not necessarily the geometry) that controls the denatured state of proteins is qualitatively quite similar to the interaction patterns derived from the statistical analysis of folded structures.

Proteins in the cellular environment encounter a multitude of interactions that do not result in the formation of complexes. The role of these nonspecific interactions is essential to the mechanisms of life and therefore actively investigated.[422] Cellular environments can be described at atomic resolution or using coarse-grained models at different levels of coarse-graining,[423] i.e. implicit solvent models, molecular-shape preserving coarse-grained models or spherical coarse-grained models of solute biomolecules. Integration of such models at different representation scales is the key challenge in the construction of integrated models, which can serve as a platform for in silico cellular models.[423]

Besides folding studies of single proteins/peptides, coarse-grained models are also applied to various specific aspects of protein folding. Example applications include studies of misfolding mechanisms;[238,272,424,425] aggregation mechanisms[26,31,424,426] (see section 4.4); the effect of post-translational modifications on protein folding and function;[427,428] the role of interplay between specific interactions (local/nonlocal, non-native/native and other) in protein folding;[408,427,429,430] or the mechanisms of chaperonin-assisted folding.[209,431−433]

4.2.2. Protein Flexibility: Small-Scale Dynamics. Time scales of biologically relevant protein fluctuations remain computationally demanding or even beyond the reach of atomistic models, and therefore coarse-grained models have emerged as an inexpensive simulation alternative.[434] One of the main problems of large scale molecular modeling is that force fields are not accurate enough, even the all-atom ones. Consequently, modeling results may be different depending on the force field choice.[21,435−437] In 2007, Orozco and colleagues examined whether different all-atom force fields provide a consistent picture of near-native dynamics in aqueous solution.[438] This comprehensive study of the most populated protein metafolds, using the four most popular force fields (OPLS, CHARMM, AMBER, and GROMOS) and explicit solvent, showed that the resulting dynamics picture is consistent among the force fields. The all-atom simulation data from the Orozco study[438] have been subsequently used as a reference for studies of different coarse-grained models as to whether they yielded comparable results.[211,439,440] These studies tested SBMs[439,440] (employing Brownian dynamics, or discrete molecular dynamics, also with a simple pseudophysical force field variant being a hybrid between the physical potential and SBM) and a knowledge-based force field derived from known folded structures (CABS model).[211] All these coarse-grained models provided a picture of near-native dynamics in good agreement with the all-atom simulation data.

Moreover, it is important to highlight that the all-atom picture of near-native protein dynamics obtained by Orozco[438] was observed on a relatively short nanosecond time scale (10 ns). Current supercomputer capabilities encompass much longer time scales and it has been shown that some important conformational transitions of a folded protein may remain undetected in submicrosecond simulations.[17,441] Future advances are expected to come with the availability of high-resolution experimental data and with developments in all-atom MD that will make this technique faster, and thus more approachable.[17] Nevertheless, in the nearest future, all-atom MD will not be suitable for flexibility analysis, and thus coarse-grained methods will continue to be the core of new efficient simulation tools.[434] Recently, several web servers using different coarse-grained modeling techniques have been proposed: CABS-flex,[210] NMSim,[385] and FlexServ[386] (see Table 3).

Functional motions of proteins can be also predicted using coarse-grained normal mode analysis (NMA).[442] Due to its success in the description of many systems and relatively simple implementation, NMA has found numerous applications in structural biology. However, certain issues about its limitations should be kept in mind, especially when large amplitude motions are considered.[256] The limitations and practicality of NMA are discussed in a review.[443]

The incorporation of protein flexibility in structure-based drug design (SBDD) is critical in many cases to obtain a valid picture of protein interaction sites.[444,445] One of the important challenges in including flexibility in SBDD studies is to choose the right level of flexibility depending on system nature and specifics of the SBDD study.[446] Therefore, future developments call for integrative methods merging SBDD approaches with experimental data and with various simulation techniques enabling efficient modeling of near-native dynamics[434,442] but also large distance movements of big macromolecules.[447]

4.3. Protein Structure Prediction

4.3.1. Importance of Computational Structure Prediction. Due to genome projects, we know a vast number of protein sequences that define their primary structure (the UniProt database currently contains around 80 million sequences and is growing exponentially[448]). Genome sequencing is now highly automated and relatively inexpensive. Experimental determination of three-dimensional (tertiary) protein structure is more challenging and still very expensive. In spite of a significant effort of many top laboratories, the number of experimentally determined structures is much smaller than the number of known sequences (PDB currently contains around 120 thousand structures[449]). This gap is still increasing. The knowledge of protein structures is one of the major requirements for understanding most of their biological functions, which is crucial for medical sciences and biotechnology. Therefore, theoretical prediction of protein native structure is one of the important challenges of molecular biology, theoretical chemistry and bioinformatics. While we can

DOI: 10.1021/acs.chemrev.6b00163
Chem. Rev. 2016, 116, 7898−7936

reliably predict a three-dimensional structure (structures) of small molecules and simple polymers, the problem of protein structure is significantly more challenging. Solvent properties and interactions with other molecules and macromolecules make the problem even more complicated and computationally demanding. Simulations of protein folding, from the denatured to the folded state based on all-atom MD, a classical simulation tool, remain generally impractical (see section 4.1). The application of coarse-grained models in protein structure prediction is therefore very appealing. Coarse-grained modeling plays an essential role not only in the ab initio prediction of protein structure (based on sequence only), but also in the most efficient strategies of comparative modeling (based on structural similarities resulting from protein homology).

4.3.2. Comparative Modeling. The comparative modeling of protein structures using their homology relations (homology modeling) is certainly one of the most successful tools of theoretical structural biology. Comparative modeling relies on the observation that proteins with similar sequences usually have similar 3D structures. This level of sequence similarity does not have to be high: even moderate sequence identity (~30%) usually implies high similarity of 3D structures, provided that sequence alignment is correct.[450] The identification of template(s) (homologous protein(s) with known 3D structure) and their sequence alignment to the target sequence is a key stage of homology modeling. It provides a crude model of the core of the target structure. The missing fragments of the modeled structure (usually, but not always, the loops connecting secondary structure elements) need to be added. Plausible structures of short fragments (up to 10 amino acids) can usually be predicted with useful accuracy and connected with the template structure. This approach is very efficiently used in the Modeller method,[351,352] a classical tool for comparative modeling based on high or moderate sequence similarity. With decreasing sequence similarity, alignment becomes less accurate and templates cover a smaller fraction of the target structure. In such cases, classical comparative modeling becomes very difficult and coarse-grained algorithms are used for modeling poorly aligned and missing fragments,[206,274,451−454] like for example in the I-TASSER automated structure prediction platform.[48]

I-TASSER is one of the most powerful tools for protein structure prediction today (see also section 4.3.4). This method is somewhat similar to the classical Modeller concept[351] in which fragments of homologous proteins are used as a core of the modeled structure. I-TASSER,[48] is based on a multilevel (multiscale modeling) approach that uses various bioinformatics and molecular modeling tools. I-TASSER uses a sophisticated sequence (target) to structure (template) three-dimensional "threading" schemes for selection of the most probable structure fragments. These fragments are used to build a core of the target structure. Subsequently a coarse-grained CAS method[173,455] (similar to CABS[100,170,173]) is used for the Monte Carlo modeling of missing or ambiguous fragments, and finally the structure is carefully refined. The method is exceptionally successful in difficult comparative modeling based on very distant homology, and thereby low sequence similarity between the target and templates used in modeling. Modifications of this method work relatively well also in more difficult structure prediction tasks, including the modeling of protein−protein complexes.[456]

Alternative to restriction of coarse-grained modeling to difficult protein fragments, different alignments and several templates could be used[457] to build a set of distance restraints for the coarse-grained modeling of the entire structure. Such a strategy may be the best choice especially for a very distant homology of template structures and it is employed for example in Phyre2[379] or CABS-fold[205] web servers for protein structure prediction.

4.3.3. Ab Initio Modeling. Using state-of-the-art tools of comparative modeling enables computational prediction of high resolution structures for a large fraction of newly sequenced proteins.[3] Nevertheless, structure prediction utilizing only the target sequence and no homology relations (termed: "ab initio" or "de novo" or "template free" modeling) is still the Holy Grail of theoretical structural biology.[47,458,459]

As mentioned in section 2.1, the first attempts at using coarse-grained models to study protein folding and to predict protein structure ab initio started about 40 years ago. Since then several similar and alternative reduced representations and sampling schemes have been used by others and in most cases it was possible to predict very low resolution structures of simple and small proteins, or protein-like systems,[170,230] peptides[460,461] or loop fragments.[206,274,451−453] Lattice models of various resolution played an important role in the early studies of protein folding.[43,462] Hierarchical schemes, which employ coarse-grained lattice models of increasing resolution, have been successfully used for de novo simulations of the protein folding process in a few small proteins.[462−464]

Presently, realistic de novo structure predictions are possible for a significant fraction of small (up to 100, or so, amino acids) and structurally not too complex proteins.[465−467] However, it has to be pointed out that the most popular contemporary algorithms for ab initio structure prediction use some general information about natural proteins. This can be done on various levels. For example, Rosetta uses sequentially similar structural fragments extracted from other proteins, not necessarily homologous.[71,378] The CABS algorithm can use secondary structure predictions and knowledge based statistical potentials derived from representative sets of protein structures. Secondary structure predictions are not necessary, although their use increases the accuracy and resolution of the resulting tertiary structures. The QUARK algorithm,[468] which proved to be most efficient in ab initio prediction in the CASP experiment[465] (see also the next section), combines the best features of Rosetta and CABS, i.e. structural fragments from known protein structures and knowledge based statistical potentials.

4.3.4. Critical Assessment of Protein Structure Prediction Methods. To validate the accuracy and efficiency of existing and newly developed methods for protein structure prediction CASP (critical assessment of methods for protein structure Prediction) experiments are organized every two years.[469] The experiment is worldwide and most research groups leading in the development of protein structure prediction methods participate (at least in some of its editions). The work scheme of CASP is the following: A couple of months before the meeting experimental structural biologists provide several sequences of proteins whose structures are near to be resolved or already resolved but not yet published. Modeling groups make theoretical predictions of the structures and deposit them on the CASP server. The organizers of the experiment collect the predictions and the experimental data on the new structures. Experts in the field evaluate the quality (accuracy) of the predicted structures. The evaluation methods are perhaps not perfect, and measurement rules of prediction

DOI: 10.1021/acs.chemrev.6b00163
Chem. Rev. 2016, 116, 7898−7936

accuracy have evolved slightly with time, but they definitely provide reasonable ranking of the accuracy of the submitted models.

Detailed discussions of successive CASP experiments can be found in dedicated issues of the Proteins journal;[469] however, some general observations are possible. First of all, the accuracy of theoretical structure prediction methods continuously (albeit rather slowly) increases. Much of this progress can be contributed to coarse-grained modeling methods, especially in the last years. The leading groups used for instance the Rosetta method,[71,378] although methods using more free-space coarse-grained sampling techniques also achieved good results, like the CABS model.[171] In some of the last CASP editions impressive predictions were provided by the Zhang group using I-TASSER (see section 4.3.2) and QUARK methods.[455,458,465]

Several other recently developed methods of multiscale modeling in structure prediction were also successful in the last CASP experiments.[470−472] Another, and less obvious, trend could also be noted. In the early CASP exercises, the best predictions were achieved by comparative modeling methods, where the key issue was the best sequence alignment and the modeling of the missing fragments was relatively simple, usually based on Modeller[351,352] (or similar) software. Difficult de novo targets were rarely predicted with realistic low-resolution accuracy. In the recent CASP experiments, the methods combining comparative modeling with tools for de novo modeling have become the most effective.

4.4. Protein Interactions

Protein−protein interactions (PPIs) play a fundamental role in controlling a wide range of biological processes including cell-to-cell interactions, signaling transduction pathways and regulatory cascades inside the cell. PPIs are responsible for protein affinity and recognition, protein−protein assembly, protein oligomerization and aggregation, and many more. Moreover, it is estimated that over 80% proteins do not operate alone but in complexes.[473] Therefore, detailed understanding of PPIs is becoming one of the major objectives of system biology. The recent development of coarse-grained techniques makes them a promising and powerful tool for PPI modeling.[29] In this section, we focus on the overview of coarse-grained methods used in modeling protein−protein or protein-peptide interactions. What's important to note, there are also lively fields of research dedicated solely to the coarse-grained modeling of specific protein interactions, for example with DNA,[202,407,474,475] RNA,[476,477] or other ligands.[478,479]

Coarse-grained models for PPI description should efficiently sample the evolution of large multiprotein systems over long time scales. On the other hand the level of coarse-graining should maintain a realistic description of side chain physicochemical properties and their interactions that control the formation of protein complexes. Moreover, modeling the protein−protein interface may require realistic prediction of induced conformational changes of interacting molecules, or even prediction of protein folding pathways. Sometimes the binding mode of two proteins may be modulated by interactions with other small molecules present in the surrounding environment.[480] Recent developments in the field of coarse-grained PPI modeling focused mainly on three areas:[29] knowledge-based molecular docking (using structural databases, bioinformatics and/or experimental data to guide the assembly of complex subunits), de novo molecular docking without the a priori localization of the binding site, and the modeling of large scale protein assembles and aggregates.

A growing amount of various experimental data provide a valuable source of information for knowledge-based molecular docking.[52] In the most favorable circumstances, the available crystal structures of protein complexes may be used as templates for the comparative structure prediction of other homologous complexes and subsequently coarse-grained techniques, often combined with all-atom scoring, can be applied for efficient refinement.[373,390,396,481,482] Such a situation is rare and in most cases very limited data are available. Fortunately, even partial structural information which allows the identification of only small fragments (or even single residues) of the interface of two interacting proteins is of great importance for PPI prediction because it largely reduces the conformational space that needs to be sampled. A variety of experimental data can be used for derivation of spatial restraints that may significantly enrich the coarse-grained modeling procedure.[483] Periole and co-workers used AMF pictures showing organization of rhodopsin in disk membranes as a starting configuration for the coarse-grained MD simulation of interacting rows of dimers.[484] Restraints derived from comparative modeling were used in the CABS coarse-grained model[72] to predict a three-dimensional structural model of a partial telomerase elongation complex composed of three essential protein domains bound to a single-stranded telomeric DNA fragment in the form of a heteroduplex.[485] The experimental data may guide the docking process by localizing protein binding interfaces or by identification of interacting conformations. Even weak distance restrains facilitated more accurate structure prediction of complex systems in the CABS method.[171] Structural restrains could be also used to validate predicted protein assemblies[486] and to improve the template-based docking procedure.[487] Guerois and co-workers proposed InterEvScore,[488] a scoring function using a coarse-grained statistical potential including two- and three-body interactions, for protein−protein docking evaluation. Combination of this potential with evolutionary information considerably improved scoring results compared to other methods (ZDOCK, ZRANK, and SPIDER) on the protein docking benchmarks tested. Due to the complexity of the PPI prediction problem the availability of experimental data that can enrich the coarse-grained modeling procedure is the decisive factor for prediction accuracy for a vast majority of protein complexes.

The de novo docking methods for PPI prediction aim at the identification of the protein−protein contact interface. This task becomes computationally extremely demanding when structural changes on docking have to be considered. Given the high computational cost of flexible docking, coarse-grained models offer an efficient and promising alternative to all-atom docking approaches. For example, Fernandez-Recio et al. developed pyDockCG,[489] a new coarse-grained potential for protein−protein docking scoring and refinement, based on the coarse-grained UNRES model.[73] In this work, new terms accounting for Coulomb electrostatics and solvation energy were proposed and tested. The pyDockCG yielded similar results to those produced by all-atom scoring function but at much lower computational cost. Another coarse-grained force field, the SCORPION,[268] was used in a series of coarse-grained MD simulations of protein−protein recognition in water for the barnase/barstar complex. The method employed coarse-grained potentials derived from the AMBER all atom force field with implementation of the new polarizable coarse-grained

DOI: 10.1021/acs.chemrev.6b00163
Chem. Rev. 2016, 116, 7898−7936

solvent (PCGS) model, whereas protein internal flexibility was accounted for by the elastic network model (ENM). The method was able to reproduce conformations very close to the native bound structures in five of a total of seven coarse-grained MD simulations. In the work of Frembgen-Kesner and Elcock,[490] application of coarse-grained Brownian dynamics (BD) simulations that included description of intermolecular hydrodynamic interactions reproduced the experimental values of association rate constants for the formation of the barnase-barstar complex. In the context of discrete molecular dynamics, a new coarse-grained force field has been recently introduced to investigate PPI and conformational sampling of multiprotein systems.[491] Interestingly, it was also reported that coarse-grained models can accurately reproduce interaction strength for protein complexes of known structure.[492]

An interesting example of a coarse-grained approach to protein docking is the ATTRACT model.[371] Starting from 2003, the ATTRACT model has been systematically improved and evolved in several docking applications. Initially, the ATTRACT force field was a molecular-shape preserving coarse-grained model with no internal main chain connectivity, with only nonbonded intermolecular terms.[371] The protein side chains were treated as an ensemble of rotamers that were discriminated during the search of relative association geometry.[371] Flexible interfacial loops, allowing for large amplitude movements, were further treated by using an ensemble of pregenerated loop conformations and a mean field approach.[493] These choices allowed very rapid search while preserving good accuracy for the search results. The initial force field was further modified for better efficiency using a different functional form for van der Waals terms.[494] Global protein flexibility was introduced along normal modes of deformation in a Gaussian network.[373] Recently, a multiscale method (ATTRACT combined with all-atom) was proposed for the refinement of protein complexes.[481] ATTRACT has been also applied to the analysis of the internal mechanics of proteins and detection of rigid amino acids using Brownian MD simulations within a Gaussian network,[495] protein−DNA docking,[496] protein-peptide docking,[381] integrative serial multi-scale modeling using interactive simulations with the Biospring engine[497,498] and to investigate large oligomeric assemblies.[498] Importantly, the ATTRACT has been successfully used for the prediction of protein complexes in different rounds of CAPRI experiments[388−390] that are discussed in the next paragraph.

The ongoing progress in protein−protein docking and PPI identification is addressed in the Critical Assessment of Prediction of Interactions (CAPRI).[499,500] CAPRI is a community-wide experiment for prediction of the molecular structure of protein complexes. From 2001 until now, 34 rounds took place. In each round, a number of protein−protein complexes whose crystal structures have been solved recently are designated as targets for blind prediction using computational methods. In various CAPRI rounds, the ATTRACT coarse-grained model (described above) proved to be successful as an ensemble docking approach that involves an implicit flexibility of protein complexes[388−390] (the classification of protein docking approaches with regard to the treatment of protein flexibility is presented in the review[445]). CAPRI results show that the main problem in docking strategies lies in the search algorithms, especially in the treatment of large-scale conformational changes, and in the scoring functions.[501−503] The coarse-grained strategies are considered promising alternatives for the future implementation of large-scale protein flexibility into on-the-fly docking (explicit flexibility) algorithms.[445,504] Such trends are represented by two methods based on Rosetta[104] and CABS[72] coarse-grained models that allow for significant structural changes during on-the-fly protein-peptide docking[213,214,383,505,506] (see Figure 8).

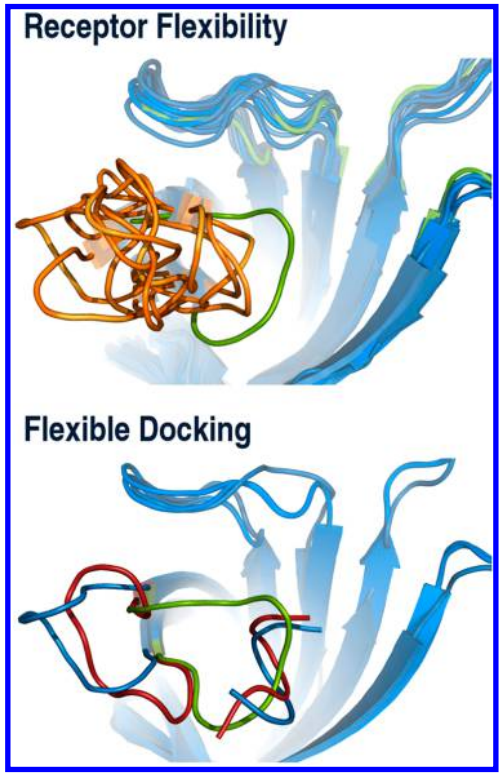


Figure 8. Protein-peptide docking with full flexibility of a protein loop region close to the binding site. The upper panel shows protein receptor flexibility during docking simulation. The starting protein structure (experimental structure in an unbound form) is shown in green. Simulated protein models are presented in blue while the flexible loop region in orange. The lower panel presents comparison of the predicted protein-peptide complex (blue) with the experimental structure of the complex (red) and the starting structure (green). Docking was performed using a coarse-grained CABS-dock method with no knowledge about the binding site or peptide conformation. During docking simulation the peptide was allowed to be fully mobile and flexible. The RMSD between the predicted and experimental peptide structure (after the best superposition of the receptor structure) was 2.03 Å. The example is described in detail elsewhere.[214]

Coarse-grained methods also remain an effective solution for an ensemble docking approach, as demonstrated by a pepATTRACT method for fully blind protein-peptide docking[381] (see Table 3).

The modeling of large scale protein assemblies and aggregates is another area in which applications of coarse-grained modeling have shown promising results.[53,89] A good example is the application of coarse-grained modeling techniques for the investigation of large complexes of membrane proteins (see section 4.5). In a recent work the coarse-grained MD approach with a push−pull-release (PPR) sampling strategy was used for a set of five well-known protein−protein complexes to describe the energy landscape and molecular forces that stabilize protein association.[507] Hansmann and co-workers[275] used the UNRES force field with MD replica exchange simulations to study the self-

DOI: 10.1021/acs.chemrev.6b00163
Chem. Rev. 2016, 116, 7898−7936

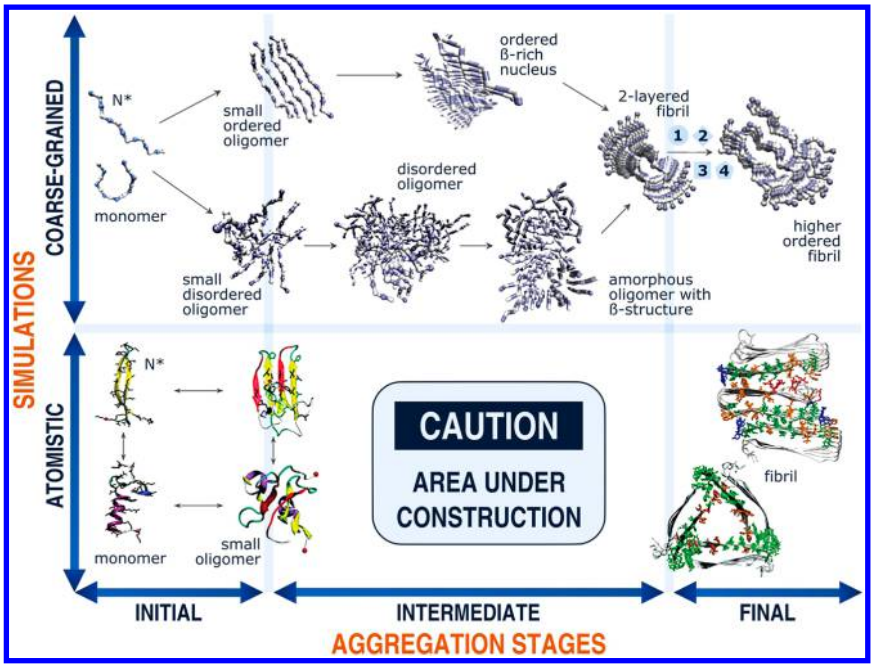


Figure 9. Comparison of all-atom and coarse-grained modeling capabilities applied in simulations of the aggregation processes. The upper panel emphasizes the ability of coarse-grained models to model a full spectrum of aggregates that occur on different aggregation pathways, from a single monomer to a highly ordered fibril, including: (1) elongation by prestructured monomer addition, (2) lateral growth by templated protofilament assembly, (3) elongation by dock-lock monomer addition, and (4) growth by dock-lock oligomer addition. The lower panel highlights the capabilities of all-atom approaches limited to the forming of small oligomers or the short-scale dynamics of formed aggregates. The "Under Construction" sign emphasize the inaccessibility of the intermediate aggregation stages to all-atom modeling. Adapted from ref.[31] Copyright 2014 American Chemical Society.

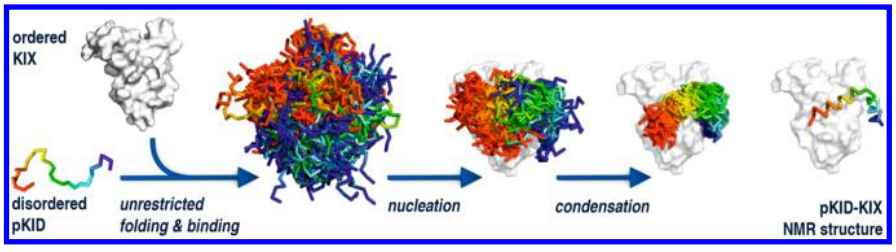


Figure 10. Mechanism of coupled folding and binding of the pKID/KIX complex as revealed by coarse-grained modeling.[212] Docking simulations allowed full flexibility of the disordered pKID during a blind search for the binding site onto the KIX surface. During docking, the movement of the KIX backbone was limited to near-native fluctuations.

assembly and PPI of a homotetrameric $\beta\beta\alpha$ (BBAT1) protein. They found that the folding and association pathway could be described by three separate steps, whereby association to a tetramer precedes and facilitates the folding of the four chains. Misfolded or partially unfolded proteins may lead to large scale aggregates, responsible for many pathological conditions. The molecular basis of aggregation-linked diseases is actively investigated using computer simulations.[251] In the reviews of Shea and co-workers[31,426] the most recent computational approaches to protein aggregation are presented, from coarse-grained models to atomistic simulations. A wide range of coarse-grained methods with different resolutions and parametrization schemes have been described.[246,508−514] Coarse-grained models can explain different aggregation pathways and cover long time scale stages of the aggregation process that are beyond the reach of atomistic simulations (see Figure 9).

Finally, a growing application area for coarse-grained modeling techniques are binding studies of intrinsically disordered proteins (IDPs)[515] or intrinsically disordered regions (IDRs) of other biomolecules.[516] There is growing evidence on IDPs or IDRs playing important functions in cellular mechanisms.[517−520] These functions are frequently involved with large-scale conformational transitions, for example from a disordered to a folded/bound state. Since the experimental characterization of IDP/IDR binding is extremely challenging,[521] simulation techniques—including coarse-grained models—have emerged as an alternative or supplementary approach.[522−525] Similarly as in the field of protein folding, the major challenge in IDP simulation is the efficient treatment of large time scale dynamics, while maintaining sufficient accuracy.[525] In the past decade, the KIX/pKID system[526] has become a model protein complex for computational studies of the folding and binding of a disordered protein. Despite a small size of the pKID/KIX complex, atomistic MD simulations are rather limited to the conformational search in the neighborhood of the native complex or to high-temperature unfolding.[23,527] Most of the computational studies of the pKID-KIX binding process used coarse-grained structure-based models (with a natively biased force field, see section 4.2.1).[528−532] Some of those studies used coarse-grained

DOI: 10.1021/acs.chemrev.6b00163
Chem. Rev. 2016, 116, 7898−7936

structure-based models extended by an additional non-native interaction component[530,531] and resulted in showing the possible important role of non-native interactions in the binding mechanism. Recently, the pKID/KIX binding mechanism has been also studied by de novo simulations (without prior information about the pKID native arrangement) using the CABS coarse-grained model.[212] Those simulations, starting from random pKID structures, yielded an ensemble of transient encounter complexes in good agreement with experimental results. The general description of the observed folding and binding mechanism is provided in Figure 10. In general, since the interest in IDP/IDR functions is relatively new, the power of coarse-graining seems not to be sufficiently exploited in the field yet. However, we have been witnessing very recently a growing number of studies on coarse-grained-based methodologies dedicated to modeling IDPs[533−535] or IDRs,[536] including studies focused on the efficient parametrization of interaction models using available experimental data,[537,538] simulation of IDP in a crowded environment,[539] or applications of coarse-grained modeling to particular IDP/IDR systems.[540−544]

4.5. Membrane Proteins

Membrane proteins (MPs) are involved in a variety of important biological functions, such as transmitting stimuli from the outside to cell interior, transport of molecules across plasma membranes and cell adhesion. MPs are also targets for over half of the currently used drugs.[545] Due to the large size of transmembrane proteins and their specifically oriented environment that needs to be somehow taken into consideration, coarse-grained modeling strategies seem to be well suited for the computational studies of MPs. The support of coarse-grained modeling is especially useful and needed in MP structure prediction, since the experimental structure determination of MPs is a very challenging task.[545,546] In this section we focus on applications of MARTINI,[116,216,217] the most popular coarse-grained model for the investigation of protein−membrane systems (see the description of MARTINI in section 2.5).

The first applications of the MARTINI model proved to be efficient during the simulation of self-assembly[547] and fusogenicity[548] of small lipid vesicles. Interestingly, the method allowed modeling rare and slowly occurring processes like flip-flop[549] and lipid desorption,[550] bending and deformation of asymmetric bilayers,[551] fusion of lipid membranes,[552] organization of proteins and peptides into lipid bilayers, protein−lipid interactions, protein oligomerization, conformational changes of tertiary protein structure and many more. The computational modeling of those phenomena requires a very large size of the modeled systems and long simulation time scales far beyond the accessible range of the current state-of-the-art atomistic simulations.

Despite the quite versatile character of the MARTINI model, it has a number of limitations that need to be taken into consideration for specific applications.[225] Due to the nature of the coarse-graining of proteins and the definition of protein topology (rather sophisticated side chain representation but a very simplified backbone) changes in protein secondary structures cannot be modeled. Moreover, accurate estimation of an effective time scale of an MD simulation depends on the particular system type and has to be considered with care.

The folding process and the structure and function of MPs are influenced by surrounding membrane environments.[553] The MARTINI model has proved to be very useful for probing protein−lipid interactions. The method was used for numerous simulations that enabled realistic prediction of binding modes in peptides and proteins to membranes, positioning proteins in lipid bilayers[554] as well as proper adaptation of membranes around proteins.[555−558] Sansom and co-workers used an extended MARTINI model for positioning a large number of proteins in lipid membranes.[559] The procedure involved a series of self-assembly MD simulations starting from systems containing protein surrounded by a random mixture of lipid and solvent molecules. In another study, multiscale simulations of a 40 amino acid C-terminal fragment of amyloid precursor protein (APP) in a DPC surfactant micelle and a POPC lipid bilayer were conducted to elucidate the role of membrane surface curvature in modulating the peptide structure.[560]

The lipids and other small molecules that make up the cell membrane may selectively bind to proteins[561] and may modulate their function. The MARTINI model was applied in numerous studies for the detection of these binding sites in MPs. Long time scale coarse-grained MARTINI MD simulations were used for the identification of the highly conserved cholesterol recognition/interaction sequence motif in the serotonin-1A receptor.[562] Other studies illustrated the binding mechanism of cholesteryl esters to cholesteryl ester transfer protein (CETP),[563] the interaction mode of heterodimeric actin-capping protein (CP) with two signaling phospholipids PA and PIP2,[564] PIP2 binding to the inwardly rectifying potassium (Kir) channel[565] and enrichment of short tail lipids near OmpA in mixtures of lipids with different tail lengths.[566] In yet another study, Arnarez and co-workers conducted an extensive set of coarse-grained MD simulations that allowed the identification of six binding sites of cardiolipin on respiratory chain complex III (cytochrome bc1, CIII).[567] A similar approach was used for the investigation of DPPC and DPPG lipid binding to the pore domain of potassium channels KcsA and chimeric KcsAKv1.3 on the structural and functional level.[568]

Studies of MP oligomerization are an important area of the application of coarse-grained protein/membrane models. The structural information on dimeric/oligomeric MPs is very important for understanding their function and mechanism of action. Sharma et al. used MARTINI to investigate the structure and assembly process of TCRα-CD3ϵ-CD3δ transmembrane domains, both in membrane and in micelle environments.[569] The T-cell receptor (TCR) together with the CD3 dimer is a key component in the primary function of T cells. MARTINI modeling of the trimeric structure allowed the identification of key interacting residues. In addition, a revised picture for the association of transmembrane domains of activating immune receptors in a membrane environment was proposed. In another study a multiscale approach employing coarse-grained MARTINI followed by all-atom MD simulations allowed the identification of the homodimer structure of two C-terminal fragments of amyloid precursor protein (C99) in the POPC bilayer and the DPC micelle.[570] Carpenter and co-workers performed coarse-grained MD simulation of the tetramerization of four transmembrane helices forming the transmembrane domain of influenza A M2 channel protein.[571] Comparison with the X-ray and NMR structures of the M2 bundle suggests that the resulting model may correspond to a closed state of the channel. Recently, the MARTINI model was applied in numerous studies describing the spontaneous self-assembly of GPCR (G-protein coupled

DOI: 10.1021/acs.chemrev.6b00163
Chem. Rev. 2016, 116, 7898−7936

receptor) dimers and oligomers in various types of cell membranes. For example Periole et al. carried out multiple self-assembly coarse-grained MD simulations of model membranes containing up to 64 molecules of the visual receptor rhodopsin over time scales reaching 100 μs.[484] The simulations allowed the identification of favored interaction interfaces between two receptors involving helices 1/8, 4/5, and 5. Furthermore, preferential interaction modes were characterized in terms of the potential of mean force (PMF) expressed as a function of interfacial distance between two receptors. A plausible picture describing the supramolecular organization of a row of dimers was also presented. In another study, Provasi and co-workers conducted extensive coarse-grained MD simulation to investigate preferred dimer interfaces of three opioid receptor subtypes: δ, κ and μ. They also addressed the possible role of interfacial lipids in modulating the rate of receptor association.[572] Coarse-grained MD simulations using the MARTINI model were also applied to assess the stability of two different dimer interfaces for β1 and β2 adrenergic receptors[573] and to explore the functional role of cholesterol concentration and its involvement in receptor organization.[480] These applications show that MARTINI, especially when combined with MD refinement, is a powerful engine for studying complex biomembrane systems.

As well as the MARTINI model or its numerous extensions, a wide range of other models have been proposed for MP simulation and modeling.[284,545,574−576] Several different coarse-grained models applying Monte Carlo simulation were used for studying the insertion of peptides into membranes[577−580] or proper description of protein interactions with lipid bilayers.[308,581] Another coarse-grained model developed by Warshel and co-workers was applied to simulate the activation process of the Kv1.2 channel.[582,583] The method was also used to analyze the energetics of translocon-assisted insertion of charged helical peptides into the membrane.[584] Recently, Feig and co-workers presented an extension of their PRIMO coarse-grained force field onto MPs.[215] The models were positively validated by comparing amino acid insertion free energy profiles with MD simulations of MPs and membrane-interacting peptides. A versatile method for modeling membrane proteins is also available in the RosettaMP framework.[585] RosettaMP allows the prediction of free energy changes upon mutation, high-resolution structural refinement, protein−protein docking and assembly of symmetric protein complexes in the membrane environment.

4.6. Integrative Modeling

Structures of large biomolecular systems are more and more often determined by integrative modeling techniques that use a combination of experimental data from various sources and different theoretical methods.[49−53] Integrative modeling approaches are also expected to provide not only static structures but also the view of conformational changes on assembly.[50,51] In comparison with the classical modeling tools of structural biology, integrative models are more complex in many aspects.[50,586] One of them is a multiscale representation requiring specific integration.[92,93] For example, the same system fragments can be represented on different levels of structural detail, and different fragments of the system can be described in different representations. Such a multiscale description can be transformed to a set of spatial restraints and provides an input to hybrid coarse-grained/all-atom methods, which are expected to provide efficient sampling and scoring.

Computational methods using coarse-grained models have already shown a great promise for a better description of protein structures, or their complexes, when combined with experimental data from NMR,[263,587−590] cryo-EM,[483,590−596] X-ray[597,598] or SAXS.[589,599−601] In particular, recent combinations of the top performing multiscale structure prediction platforms, Rosetta and I-TASSER, with experimental measurements resulted in spectacular prediction results. For example, the refinement of protein NMR structures using Rosetta with experimental NMR restraints yielded more accurate structures than corresponding X-ray crystal structures.[263] Another interesting example is the performance of the NMR-I-TASSER method (an adaptation of the I-TASSER platform) in the recent critical assessment of automated structure determination of proteins from NMR Data (CASP-NMR) experiment.[590] It was shown that even using only the coarse-grained conformational search, NMR-I-TASSER can consistently produce good resolution models (<2 Å). This makes NMR-I-TASSER very promising for applications in combination with the classical all-atom refinement tools, in particular for the efficient structure determination of large proteins.

5. CONCLUDING REMARKS

Most biomolecular systems, including proteins and their complexes, are too complicated to be efficiently handled by classical molecular modeling tools. This is caused not only by the large molecular size but also by the time-scale of important processes and specific interaction patterns. The coarse-grained modeling approaches outlined in this review seem to be the computational methods of choice in solving many fundamental problems of theoretical and practical molecular biology and medicinal chemistry. When using or designing coarse-grained-based modeling methods, it is necessary to consider several important choices, such as the resolution level and specific design of coarse-grained representation, model of interactions, sampling schemes, and finally the efficient use of experimental (and theoretical) data and effective connection of coarse-grained computations with atom level simulations.

When designing coarse-grained modeling methods, the representation of atomistic structures on a coarse-grained level requires precise definition. The choice of representation determines to a large extent the possible options of force field and sampling, i.e., the compromise between accuracy and computational efficiency.[87,407] The smaller the number of explicitly treated united atoms (or pseudoatoms) representing fragments of protein chains, the faster simulation, and the lower accuracy. Very efficient models based on three/four united atoms per amino acid residue accelerate simulations by 3−4 orders of magnitude in comparison with classical all-atom MD simulations.[211,271] Nevertheless, it is useful to develop even more simplified models dedicated to large protein systems with realistic connection with all-atom resolution schemes.[28,225,314,407] On the other hand, some applications require a fine level of coarse-graining.[105,602] For example, the coarse-grained modeling of solvent and protein interaction effects with small molecules is very challenging[603−605] and it very likely will be an important subject of future research. Dedicated hardware and specific programming may provide additional speedup factors.[606−608]

Improvement of interaction models is of primary importance in the protein modeling field. Current force fields are not accurate enough, and it is possible to obtain different results using different force fields, even the all-atom ones.[21,435−437]

DOI: 10.1021/acs.chemrev.6b00163
Chem. Rev. 2016, 116, 7898−7936

The most straightforward, although definitely not trivial, interaction schemes for coarse-grained models could be derived from the atom-level force fields of classical molecular mechanics.[28,30,609] It is not easy to ensure "transferability" of such "physics-based" interaction schemes between different systems or different environments, although significant progress has been achieved recently.[30,610] Alternatively to the "physics-based" approach it is possible to derive statistical "knowledge-based" force fields that generalize coarse-grained structural regularities seen in the known protein structures. In the past few years modeling schemes based on statistical force fields have proven to be most successful in protein structure prediction (I-TASSER,[48] Rosetta,[104] and CABS[205]). These methods use various combinations of statistical potentials in which sequence specific interactions are based on the statistical analysis of structural data, sometimes limited to sequentially analogous or homologous proteins. Surprisingly, models employing "knowledge-based" interactions derived from regularities observed in static experimental structures seem to provide quite realistic pictures of protein folding pathways and protein dynamics. Nevertheless, the problem of transferability, or rather the balance between specificity and accuracy of statistical potentials (for example between force fields for globular proteins versus force fields for membrane proteins, or for protein monomers and for protein−protein complexes, etc.) needs further studies. Finally, any of the interaction schemes of coarse-grained models could be combined with "restrains" derived from various, often fragmentary, experimental data, enabling their deeper interpretation.

Various sampling schemes, including Molecular Dynamics, Monte Carlo, or their combinations can be applied to coarse-grained models.[181−185] Coarse-grained representation enables (and in some sense enforces) significantly smoother energy curves for interaction schemes. Therefore, the apparent time step in coarse-grained MD[269] could be broader than required for atom-level simulations. This provides additional speedup of coarse-grained simulations. Very efficient Monte Carlo sampling schemes could also be used, and these are the natural choice for discrete models. Properly designed MC algorithms based on local conformational modifications can provide quite realistic pictures of long time dynamics. The replica-exchange and other multicopy schemes of MD (or MC) simulations could also be very useful in studies of coarse-grained models of biomacromolecules.[183,184] Recent progress in sampling has also brought another "Big Data" challenge: how to merge and analyze the massive amounts of simulation data.[199−201]

The multiscale or integrative strategies of molecular modeling have been rapidly developing in the last years. The efficient use of coarse-grained models usually requires rigorous reconstruction of the atom-level representation. This is not a trivial task, and satisfactory solutions exist only for well-studied moderate resolution coarse-grained models.[101,102,360] Fast methods of dependable transitions between various levels of representation are needed, and probably require designing specific statistical potentials to ensure not only realistic spatial fidelity but also a reasonable switch between dynamics profiles. The existing coarse-grained modeling tools dedicated to particular types of macromolecules (e.g., lipids,[228,280] DNA,[475,611] RNA,[612,613] or carbohydrates[614,615]) require integration, or better integration, with coarse-grained protein models and within well-defined multiscale modeling schemes.

Over the last decades, we have learned about the power of coarse-grained molecular modeling of proteins and their complexes,[19−32] in particular when combined with higher resolution models.[25,37−40] The application of coarse-grained modeling in combination with all-atom refinement tools, bioinformatics and fragmentary experimental data already plays a crucial role in protein structure prediction.[48,104,379] Similar progress is now being made in predicting the structure of protein complexes.[52,55] The efficient multiscale modeling of long time biomolecular dynamics[50,51] is the next challenge. A growing number of real-time measurements of how cell components (proteins, nucleic acids, lipids, and others) perform their functions is expected to be a further stimulus to construct new coarse-grained modeling tools, including whole-cell models.[54,55,423]

AUTHOR INFORMATION

Corresponding Author

*E-mail: kolinski@chem.uw.edu.pl. Phone: +48 22 55 26 365. Fax: (+48) 22 55 26 364.

Notes

The authors declare no competing financial interest.

Biographies

Sebastian Kmiecik received his Ph.D. in 2007 from the Faculty of Chemistry, University of Warsaw, Poland. In 2007, he joined a biotechnology startup company, Selvita, where as a Project Manager and later the Head of Modeling, he helped to develop the company's R&D programs and services. In 2010, he again joined the Faculty of Chemistry, University of Warsaw. Since that time his academic work has been focused on the development of multiscale methods for protein modeling and their application in studies of protein function. In 2015, he received a D.Sc. degree.

Dominik Gront received his Ph.D. in 2006 from the Faculty of Chemistry, University of Warsaw. After postdoctoral training at the University of Virginia, Charlottesville, VA and University of Washington, Seattle, WA, he rejoined the Faculty of Chemistry, University of Warsaw. In his research he focuses on novel coarse-grained methods for biomacromolecular modeling. He actively develops Rosetta and BioShell software packages.

Michal Kolinski received his Ph.D. in 2010 from the Institute of Biochemistry and Biophysics, Polish Academy of Sciences in Warsaw. After that, he became a postdoc in the Laboratory of Medicinal Chemistry and Neuroengineering at the Medical University of Lublin, Poland. In 2011, he joined Bioinformatics Laboratory in the Mossakowski Medical Research Centre, Polish Academy of Sciences in Warsaw. Currently he is an Assistant Professor and his research interests include the application of computational methods for prediction of membrane protein structure, function, and ligand binding.

Lukasz Wieteska received his Master's degree from the Faculty of Pharmacy, Medical University of Lodz in 2009 and Bachelor's degree from the Faculty of Mathematics and Computer Science, University of Lodz in 2011. In 2014, he defended his Ph.D. thesis at the Department of Medical Biochemistry, Medical University of Lodz. Currently he is a Postdoc in the Laboratory of Theory of Biopolymers, University of Warsaw. His scientific interests include application of computational methods in biophysical studies of proteins and protein modifications relevant to medicine and biotechnology.

Aleksandra Dawid graduated from the Faculty of Chemistry, University of Warsaw, Poland in 2014. She is now a Ph.D. student at the Department of Chemistry, working in the Laboratory of Theory of Biopolymers on the development of knowledge-based statistical

DOI: 10.1021/acs.chemrev.6b00163
Chem. Rev. 2016, 116, 7898−7936

force fields for simulations of large biomacromolecular systems. She is also developing a new low resolution coarse-grained model for the long time multiscale modeling of proteins.

Andrzej Kolinski studied chemistry at the University of Warsaw and completed his Ph.D. thesis on the computer modeling of polymerization processes at the Department of Chemistry, University of Warsaw in 1979. Then he was directly appointed by the same department first as an assistant professor and then later as a full professor. Between 1985 and 2005 he was also part time appointed at various research institutions in the U.S.A., mostly at the Department of Molecular Biology, Scripps Research Institute (San Diego, California), but also for shorter periods at the Department of Chemistry, Washington University (Saint Louis), at Donald Danforth Plant Science Center (Saint Louis), and at the Center of Excellence in Bioinformatics (Buffalo). Since 1998 he is the head of the Laboratory of Theory of Biopolymers at the Faculty of Chemistry, University of Warsaw. His current research interests include structural bioinformatics, development of a new modeling tool for simulations of complex biomacromolecules, and computer-aided rational drug design.

ACKNOWLEDGMENTS

S.K., D.G., L.W., A.D., and A.K. acknowledge support from the National Science Center (NCN, Poland) Grant [MAESTRO2014/14/A/ST6/00088]. M.K. acknowledges support from the Polish Ministry of Science and Higher Education Grant [IP2012 016372].

REFERENCES

(1) De Lucrezia, D.; Slanzi, D.; Poli, I.; Polticelli, F.; Minervini, G. Do natural proteins differ from random sequences polypeptides? Natural vs. random proteins classification using an evolutionary neural network. *PLoS One* **2012**, *7*, e36634.
(2) UniProt: a hub for protein information. *Nucleic Acids Res.* **2015**, *43*, D204−212.10.1093/nar/gku989
(3) Haas, J.; Roth, S.; Arnold, K.; Kiefer, F.; Schmidt, T.; Bordoli, L.; Schwede, T. The Protein Model Portal–a comprehensive resource for protein structure and model information. *Database* **2013**, *2013*, bat031.
(4) Bu, Z.; Callaway, D. J. Proteins move! Protein dynamics and long-range allostery in cell signaling. *Adv. Protein Chem. Struct. Biol.* **2011**, *83*, 163−221.
(5) Chouard, T. Structural biology: Breaking the protein rules. *Nature* **2011**, *471*, 151−153.
(6) Henzler-Wildman, K.; Kern, D. Dynamic personalities of proteins. *Nature* **2007**, *450*, 964−972.
(7) Schmidt, T.; Bergner, A.; Schwede, T. Modelling three-dimensional protein structures for applications in drug design. *Drug Discovery Today* **2014**, *19*, 890−897.
(8) Schwede, T.; Sali, A.; Honig, B.; Levitt, M.; Berman, H. M.; Jones, D.; Brenner, S. E.; Burley, S. K.; Das, R.; Dokholyan, N. V.; et al. Outcome of a workshop on applications of protein models in biomedical research. *Structure* **2009**, *17*, 151−159.
(9) Khoury, G. A.; Smadbeck, J.; Kieslich, C. A.; Floudas, C. A. Protein folding and de novo protein design for biotechnological applications. *Trends Biotechnol.* **2014**, *32*, 99−109.
(10) Lounnas, V.; Ritschel, T.; Kelder, J.; McGuire, R.; Bywater, R. P.; Foloppe, N. Current progress in Structure-Based Rational Drug Design marks a new mindset in drug discovery. *Comput. Struct. Biotechnol. J.* **2013**, *5*, e201302011.
(11) Ferreira, L. G.; Dos Santos, R. N.; Oliva, G.; Andricopulo, A. D. Molecular docking and structure-based drug design strategies. *Molecules* **2015**, *20*, 13384−13421.
(12) Wang, T.; Wu, M. B.; Zhang, R. H.; Chen, Z. J.; Hua, C.; Lin, J. P.; Yang, L. R. Advances in Computational Structure-Based Drug Design and Application in Drug Discovery. *Curr. Top. Med. Chem.* **2015**, *16*, 901−916.
(13) Ozboyaci, M.; Kokh, D. B.; Corni, S.; Wade, R. C. Modeling and simulation of protein-surface interactions: achievements and challenges. *Q. Rev. Biophys.* **2016**, *49*, e4.
(14) Vendruscolo, M.; Dobson, C. M. Protein dynamics: Moore's law in molecular biology. *Curr. Biol.* **2011**, *21*, R68−70.
(15) Shaw, D. E.; Grossman, J. P.; Bank, J. A.; Batson, B.; Butts, J. A.; Chao, J. C.; Deneroff, M. M.; Dror, R. O.; Even, A.; Fenton, C. H. et al. (2014) Anton 2: raising the bar for performance and programmability in a special-purpose molecular dynamics supercomputer. In Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis; IEEE Press: New Orleans; pp 41−53.
(16) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. How fast-folding proteins fold. *Science* **2011**, *334*, 517−520.
(17) Piana, S.; Klepeis, J. L.; Shaw, D. E. Assessing the accuracy of physical models used in protein-folding simulations: quantitative evidence from long molecular dynamics simulations. *Curr. Opin. Struct. Biol.* **2014**, *24*, 98−105.
(18) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. Atomistic description of the folding of a dimeric protein. *J. Phys. Chem. B* **2013**, *117*, 12935−12942.
(19) Kolinski, A.; Skolnick, J. Reduced models of proteins and their applications. *Polymer* **2004**, *45*, 511−524.
(20) Tozzini, V. Coarse-grained models for proteins. *Curr. Opin. Struct. Biol.* **2005**, *15*, 144−150.
(21) Scheraga, H. A.; Khalili, M.; Liwo, A. Protein-folding dynamics: overview of molecular simulation techniques. *Annu. Rev. Phys. Chem.* **2007**, *58*, 57−83.
(22) Clementi, C. Coarse-grained models of protein folding: toy models or predictive tools? *Curr. Opin. Struct. Biol.* **2008**, *18*, 10−15.
(23) Espinoza-Fonseca, L. M. Thermodynamic aspects of coupled binding and folding of an intrinsically disordered protein: a computational alanine scanning study. *Biochemistry* **2009**, *48*, 11332−11334.
(24) Rader, A. J. Coarse-grained models: getting more with less. *Curr. Opin. Pharmacol.* **2010**, *10*, 753−759.
(25) Kamerlin, S. C.; 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.
(26) Wu, C.; Shea, J. E. Coarse-grained models for protein aggregation. *Curr. Opin. Struct. Biol.* **2011**, *21*, 209−220.
(27) Takada, S. Coarse-grained molecular simulations of large biomolecules. *Curr. Opin. Struct. Biol.* **2012**, *22*, 130−137.
(28) Saunders, M. G.; Voth, G. A. Coarse-graining methods for computational biology. *Annu. Rev. Biophys.* **2013**, *42*, 73−93.
(29) Baaden, M.; Marrink, S. J. Coarse-grain modelling of protein-protein interactions. *Curr. Opin. Struct. Biol.* **2013**, *23*, 878−886.
(30) Kar, P.; Feig, M. Recent advances in transferable coarse-grained modeling of proteins. *Adv. Protein Chem. Struct. Biol.* **2014**, *96*, 143−180.
(31) Morriss-Andrews, A.; Shea, J. E. Simulations of Protein Aggregation: Insights from Atomistic and Coarse-Grained Models. *J. Phys. Chem. Lett.* **2014**, *5*, 1899−1908.
(32) Ingolfsson, H. I.; Lopez, C. A.; Uusitalo, J. J.; de Jong, D. H.; Gopal, S. M.; Periole, X.; Marrink, S. J. The power of coarse graining in biomolecular simulations. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4*, 225−248.
(33) Voth, G. A. *Coarse-Graining of Condensed Phase and Biomolecular Systems*; CRC Press: Boca Raton, FL, 2008; p 456.
(34) Levitt, M.; Warshel, A. Computer-Simulation of Protein Folding. *Nature* **1975**, *253*, 694−698.
(35) Warshel, A.; Levitt, M. Theoretical Studies of Enzymic Reactions - Dielectric, Electrostatic and Steric Stabilization of Carbonium-Ion in Reaction of Lysozyme. *J. Mol. Biol.* **1976**, *103*, 227−249.

DOI: 10.1021/acs.chemrev.6b00163
Chem. Rev. 2016, 116, 7898−7936

(36) Levitt, M. Birth and future of multiscale modeling for macromolecular systems (Nobel Lecture). *Angew. Chem., Int. Ed.* **2014**, *53*, 10006−10018.
(37) Ayton, G. S.; Noid, W. G.; Voth, G. A. Multiscale modeling of biomolecular systems: in serial and in parallel. *Curr. Opin. Struct. Biol.* **2007**, *17*, 192−198.
(38) Tozzini, V. Multiscale modeling of proteins. *Acc. Chem. Res.* **2010**, *43*, 220−230.
(39) Kolinski, A. *Multiscale Approaches to Protein Modeling*; Springer: New York, 2011; p 355.
(40) Zhou, H. X. Theoretical frameworks for multiscale modeling and simulation. *Curr. Opin. Struct. Biol.* **2014**, *25*, 67−76.
(41) Sippl, M. J. Knowledge-based potentials for proteins. *Curr. Opin. Struct. Biol.* **1995**, *5*, 229−235.
(42) Carlsen, M.; Koehl, P.; Rogen, P. On the importance of the distance measures used to train and test knowledge-based potentials for proteins. *PLoS One* **2014**, *9*, e109335.
(43) Kolinski, A.; Skolnick, J. *Lattice models of protein folding, dynamics, and thermodynamics*; R. G. Landes: Austin, TX, 1996.
(44) Mirny, L.; Shakhnovich, E. Protein folding theory: from lattice to all-atom models. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 361−396.
(45) Blaszczyk, M.; Gront, D.; Kmiecik, S.; Ziolkowska, K.; Panek, M.; Kolinski, A. In *Computational Methods to Study the Structure and Dynamics of Biomolecules and Biomolecular Processes: From Bioinformatics to Molecular Quantum Mechanics*; Liwo, A., Ed.; Springer: Berlin, 2014.
(46) Park, H.; DiMaio, F.; Baker, D. CASP11 refinement experiments with ROSETTA. *Proteins: Struct., Funct., Genet.* **2015**, n/a.
(47) Ovchinnikov, S.; Kim, D. E.; Wang, R. Y.; Liu, Y.; DiMaio, F.; Baker, D. Improved de novo structure prediction in CASP11 by incorporating Co-evolution information into rosetta. *Proteins: Struct., Funct., Genet.* **2015**, n/a.
(48) Yang, J. Y.; Yan, R. X.; Roy, A.; Xu, D.; Poisson, J.; Zhang, Y. The I-TASSER Suite: protein structure and function prediction. *Nat. Methods* **2014**, *12*, 7−8.
(49) Webb, B.; Lasker, K.; Velazquez-Muriel, J.; Schneidman-Duhovny, D.; Pellarin, R.; Bonomi, M.; Greenberg, C.; Raveh, B.; Tjioe, E.; Russel, D.; et al. Modeling of proteins and their assemblies with the Integrative Modeling Platform. *Methods Mol. Biol.* **2014**, *1091*, 277−295.
(50) Sali, A.; Berman, H. M.; Schwede, T.; Trewhella, J.; Kleywegt, G.; Burley, S. K.; Markley, J.; Nakamura, H.; Adams, P.; Bonvin, A. M.; et al. Outcome of the First wwPDB Hybrid/Integrative Methods Task Force Workshop. *Structure* **2015**, *23*, 1156−1167.
(51) Tamo, G. E.; Abriata, L. A.; Dal Peraro, M. The importance of dynamics in integrative modeling of supramolecular assemblies. *Curr. Opin. Struct. Biol.* **2015**, *31*, 28−34.
(52) Rodrigues, J. P.; Bonvin, A. M. Integrative computational modeling of protein interactions. *FEBS J.* **2014**, *281*, 1988−2003.
(53) Reddy, T.; Sansom, M. S. Computational virology: From the inside out. *Biochim. Biophys. Acta, Biomembr.* **2016**, *1858*, 1610.
(54) Hyeon, C.; Thirumalai, D. Capturing the essence of folding and functions of biomolecules using coarse-grained models. *Nat. Commun.* **2011**, *2*, 487.
(55) Spiga, E.; Degiacomi, M. T.; Dal Peraro, M. New strategies for integrative dynamic modeling of macromolecular assembly. *Adv. Protein Chem. Struct. Biol.* **2014**, *96*, 77−111.
(56) Levinthal, C. Are There Pathways for Protein Folding. *J. Chim. Phys. Pcb.* **1968**, *65*, 44.
(57) Levitt, M. A simplified representation of protein conformations for rapid simulation of protein folding. *J. Mol. Biol.* **1976**, *104*, 59−107.
(58) Hagler, A. T.; Honig, B. On the formation of protein tertiary structure on a computer. *Proc. Natl. Acad. Sci. U. S. A.* **1978**, *75*, 554−558.
(59) Wilson, C.; Doniach, S. A computer model to dynamically simulate protein folding: studies with crambin. *Proteins: Struct., Funct., Genet.* **1989**, *6*, 193−209.
(60) Crippen, G. M.; Ponnuswamy, P. K. Determination of an empirical energy function for protein conformational analysis by energy embedding. *J. Comput. Chem.* **1987**, *8*, 972−981.
(61) Miyazawa, S.; Jernigan, R. L. Estimation of effective inter-residue contact energies from protein crystal structures: quasi-chemical approximation. *Macromolecules* **1985**, *18*, 534−552.
(62) Sun, S. Reduced representation model of protein structure prediction: statistical potential and genetic algorithms. *Protein Sci.* **1993**, *2*, 762−785.
(63) Wallqvist, A.; Ullner, M. A simplified amino acid potential for use in structure predictions of proteins. *Proteins: Struct., Funct., Genet.* **1994**, *18*, 267−280.
(64) Honeycutt, J. D.; Thirumalai, D. Metastability of the Folded States of Globular-Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **1990**, *87*, 3526−3529.
(65) Covell, D. G. Folding Protein Alpha-Carbon Chains into Compact Forms by Monte-Carlo Methods. *Proteins: Struct., Funct., Genet.* **1992**, *14*, 409−420.
(66) Krigbaum, W. R.; Lin, S. F. Monte-Carlo Simulation of Protein Folding Using a Lattice Model. *Macromolecules* **1982**, *15*, 1135−1145.
(67) Skolnick, J.; Kolinski, A. Simulations of the folding of a globular protein. *Science* **1990**, *250*, 1121−1125.
(68) Hinds, D. A.; Levitt, M. A Lattice Model for Protein-Structure Prediction at Low Resolution. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 2536−2540.
(69) Taketomi, H.; Kano, F.; Go, N. The effect of amino acid substitution on protein-folding and -unfolding transition studied by computer simulation. *Biopolymers* **1988**, *27*, 527−559.
(70) Voegler Smith, A.; Hall, C. K. alpha-helix formation: discontinuous molecular dynamics on an intermediate-resolution protein model. *Proteins: Struct., Funct., Genet.* **2001**, *44*, 344−360.
(71) Rohl, C.; Strauss, C.; Misura, K.; Baker, D. In *Numerical Computer Methods, Part D*; Elsevier: Amsterdam, 2004; Vol. *383*.
(72) Kolinski, A. Protein modeling and structure prediction with a reduced representation. *Acta Biochim. Polym.* **2004**, *51*, 349−371.
(73) Liwo, A.; Baranowski, M.; Czaplewski, C.; Golas, E.; He, Y.; Jagiela, D.; Krupa, P.; Maciejczyk, M.; Makowski, M.; Mozolewska, M. A.; et al. A unified coarse-grained model of biological macromolecules based on mean-field multipole-multipole interactions. *J. Mol. Model.* **2014**, *20*, 2306.
(74) Kolinski, A.; Jaroszewski, L.; Rotkiewicz, P.; Skolnick, J. An efficient Monte Carlo model of protein chains. Modeling the short-range correlations between side group centers of mass. *J. Phys. Chem. B* **1998**, *102*, 4628−4637.
(75) Shakhnovich, E. Protein folding thermodynamics and dynamics: where physics, chemistry, and biology meet. *Chem. Rev.* **2006**, *106*, 1559−1588.
(76) Dill, K. A.; MacCallum, J. L. The protein-folding problem, 50 years on. *Science* **2012**, *338*, 1042−1046.
(77) Yue, K.; Fiebig, K. M.; Thomas, P. D.; Chan, H. S.; Shakhnovich, E. I.; Dill, K. A. A test of lattice protein folding algorithms. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 325−329.
(78) Dill, K. A.; Bromberg, S.; Yue, K.; Fiebig, K. M.; Yee, D. P.; Thomas, P. D.; Chan, H. S. Principles of protein folding–a perspective from simple exact models. *Protein Sci.* **1995**, *4*, 561−602.
(79) Shakhnovich, E. I.; Gutin, A. M. Formation of Unique Structure in Polypeptide-Chains Theoretical Investigation with the Aid of a Replica Approach. *Biophys. Chem.* **1989**, *34*, 187−199.
(80) Bechini, A. On the Characterization and Software Implementation of General Protein Lattice Models. *PLoS One* **2013**, *8*.e5950410.1371/journal.pone.0059504
(81) Skolnick, J.; Kolinski, A.; Yaris, R. Monte Carlo simulations of the folding of beta-barrel globular proteins. *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85*, 5057−5061.
(82) Skolnick, J.; Kolinski, A. Dynamic Monte Carlo simulations of a new lattice model of globular protein folding, structure and dynamics. *J. Mol. Biol.* **1991**, *221*, 499−531.

DOI: 10.1021/acs.chemrev.6b00163
Chem. Rev. 2016, 116, 7898−7936

(83) Hao, M. H.; Scheraga, H. A. Monte-Carlo Simulation of a First-Order Transition for Protein-Folding. *J. Phys. Chem.* **1994**, *98*, 4940−4948.
(84) Park, B. H.; Levitt, M. The Complexity and Accuracy of Discrete State Models of Protein-Structure. *J. Mol. Biol.* **1995**, *249*, 493−507.
(85) Godzik, A.; Kolinski, A.; Skolnick, J. Lattice Representations of Globular-Proteins - How Good Are They. *J. Comput. Chem.* **1993**, *14*, 1194−1202.
(86) Alexander, P. A.; He, Y.; Chen, Y.; Orban, J.; Bryan, P. N. A minimal sequence code for switching protein structure and function. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 21149−21154.
(87) Riniker, S.; Allison, J. R.; van Gunsteren, W. F. On developing coarse-grained models for biomolecular simulation: a review. *Phys. Chem. Chem. Phys.* **2012**, *14*, 12423−12430.
(88) Trylska, J. Coarse-grained models to study dynamics of nanoscale biomolecules and their applications to the ribosome. *J. Phys.: Condens. Matter* **2010**, *22*, 453101.
(89) Saunders, M. G.; Voth, G. A. Coarse-graining of multiprotein assemblies. *Curr. Opin. Struct. Biol.* **2012**, *22*, 144−150.
(90) Mills, Z. G.; Mao, W.; Alexeev, A. Mesoscale modeling: solving complex flows in biology and biotechnology. *Trends Biotechnol.* **2013**, *31*, 426−434.
(91) Coulon, A.; Beslon, G.; Gandrillon, O. Large multiprotein structures modeling and simulation: the need for mesoscopic models. *Methods Mol. Biol.* **2008**, *484*, 537−558.
(92) Dama, J. F.; Sinitskiy, A. V.; McCullagh, M.; Weare, J.; Roux, B.; Dinner, A. R.; Voth, G. A. The Theory of Ultra-Coarse-Graining. 1. General Principles. *J. Chem. Theory Comput.* **2013**, *9*, 2466−2480.
(93) Davtyan, A.; Dama, J. F.; Sinitskiy, A. V.; Voth, G. A. The Theory of Ultra-Coarse-Graining. 2. Numerical Implementation. *J. Chem. Theory Comput.* **2014**, *10*, 5265−5275.
(94) Chen, J.; Xie, Z. R.; Wu, Y. Elucidating the general principles of cell adhesion with a coarse-grained simulation model. *Mol. BioSyst.* **2016**, *12*, 205−218.
(95) Kolinski, A.; Skolnick, J. Assembly of protein structure from sparse experimental data: An efficient Monte Carlo model. *Proteins: Struct., Funct., Genet.* **1998**, *32*, 475−494.
(96) Kolinski, A.; Betancourt, M. R.; Kihara, D.; Rotkiewicz, P.; Skolnick, J. Generalized comparative modeling (GENECOMP): a combination of sequence comparison, threading, and lattice modeling for protein structure prediction and refinement. *Proteins: Struct., Funct., Genet.* **2001**, *44*, 133−149.
(97) Kolinski, A.; Ilkowski, B.; Skolnick, J. Dynamics and thermodynamics of beta-hairpin assembly: insights from various simulation techniques. *Biophys. J.* **1999**, *77*, 2942−2952.
(98) Boniecki, M.; Rotkiewicz, P.; Skolnick, J.; Kolinski, A. Protein fragment reconstruction using various modeling techniques. *J. Comput.-Aided Mol. Des.* **2003**, *17*, 725−738.
(99) Stumpff-Kane, A. W.; Maksimiak, K.; Lee, M. S.; Feig, M. Sampling of near-native protein conformations during protein structure refinement using a coarse-grained model, normal modes, and molecular dynamics simulations. *Proteins: Struct., Funct., Genet.* **2008**, *70*, 1345−1356.
(100) Kihara, D.; Zhang, Y.; Lu, H.; Kolinski, A.; Skolnick, J. Ab initio protein structure prediction on a genomic scale: application to the Mycoplasma genitalium genome. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 5993−5998.
(101) Kmiecik, S.; Gront, D.; Kouza, M.; Kolinski, A. From coarse-grained to atomic-level characterization of protein dynamics: transition state for the folding of B domain of protein A. *J. Phys. Chem. B* **2012**, *116*, 7026−7032.
(102) Xu, D.; Zhang, Y. Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. *Biophys. J.* **2011**, *101*, 2525−2534.
(103) Kazmierkiewicz, R.; Liwo, A.; Scheraga, H. A. Addition of side chains to a known backbone with defined side-chain centroids. *Biophys. Chem.* **2002**, *100*, 261−280.
(104) Das, R.; Baker, D. Macromolecular modeling with rosetta. *Annu. Rev. Biochem.* **2008**, *77*, 363−382.
(105) Gopal, S. M.; Mukherjee, S.; Cheng, Y. M.; Feig, M. PRIMO/PRIMONA: a coarse-grained model for proteins and nucleic acids that preserves near-atomistic accuracy. *Proteins: Struct., Funct., Genet.* **2010**, *78*, 1266−1281.
(106) Kar, P.; Gopal, S. M.; Cheng, Y. M.; Predeus, A.; Feig, M. PRIMO: A Transferable Coarse-grained Force Field for Proteins. *J. Chem. Theory Comput.* **2013**, *9*, 3769−3788.
(107) Skolnick, J. In quest of an empirical potential for protein structure prediction. *Curr. Opin. Struct. Biol.* **2006**, *16*, 166−171.
(108) Vanommeslaeghe, K.; MacKerell, A. D., Jr. CHARMM additive and polarizable force fields for biophysics and computer-aided drug design. *Biochim. Biophys. Acta, Gen. Subj.* **2015**, *1850*, 861−871.
(109) Oostenbrink, C.; Villa, A.; Mark, A.; van Gunsteren, W. 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.
(110) Fain, B.; Xia, Y.; Levitt, M. Design of an optimal Chebyshev-expanded discrimination function for globular proteins. *Protein Sci.* **2002**, *11*, 2010−2021.
(111) Basdevant, N.; Borgis, D.; Ha-Duong, T. A coarse-grained protein-protein potential derived from an all-atom force field. *J. Phys. Chem. B* **2007**, *111*, 9390−9399.
(112) Cascella, M.; Vanni, S. In *Chemical Modelling: Vol. 12*; The Royal Society of Chemistry: London, 2016; Vol. *12*.
(113) Brooks, B.; Bruccoleri, R.; Olafson, B.; States, D.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187−217.
(114) Lee, S.; Tran, A.; Allsopp, M.; Lim, J. B.; Henin, J.; Klauda, J. B. CHARMM36 united atom chain model for lipids and surfactants. *J. Phys. Chem. B* **2014**, *118*, 547−556.
(115) Schmid, N.; Eichenberger, A. P.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A. E.; van Gunsteren, W. F. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur. Biophys. J.* **2011**, *40*, 843−856.
(116) de Jong, D. H.; Singh, G.; Bennett, W. F. D.; Arnarez, C.; Wassenaar, T. A.; Schafer, L. V.; Periole, X.; Tieleman, D. P.; Marrink, S. J. Improved Parameters for the Martini Coarse-Grained Protein Force Field. *J. Chem. Theory Comput.* **2013**, *9*, 687−697.
(117) Izvekov, S.; Parrinello, M.; Burnham, C.; Voth, G. Effective force fields for condensed phase systems from ab initio molecular dynamics simulation: A new method for force-matching. *J. Chem. Phys.* **2004**, *120*, 10896−10913.
(118) Ercolessi, F.; Adams, J. B. Interatomic Potentials from First-Principles Calculations: The Force-Matching Method. *Europhys. Lett.* **1994**, *26*, 583.
(119) Schommers, W. Pair potentials in disordered many-particle systems: A study for liquid gallium. *Phys. Rev. A: At., Mol., Opt. Phys.* **1983**, *28*, 3599−3605.
(120) Henderson, R. L. A uniqueness theorem for fluid pair correlation functions. *Phys. Lett. A* **1974**, *49*, 197−198.
(121) Reith, D.; Pütz, M.; Müller-Plathe, F. Deriving effective mesoscale potentials from atomistic simulations. *J. Comput. Chem.* **2003**, *24*, 1624−1636.
(122) Moore, T. C.; Iacovella, C. R.; McCabe, C. Derivation of coarse-grained potentials via multistate iterative Boltzmann inversion. *J. Chem. Phys.* **2014**, *140*, 224104.
(123) Lyubartsev, A.; Laaksonen, A. Calculation of effective interaction potentials from radial distribution functions: A reverse Monte Carlo approach. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **1995**, *52*, 3730−3737.
(124) Lyubartsev, A.; Mirzoev, A.; Chen, L.; Laaksonen, A. Systematic coarse-graining of molecular models by the Newton inversion method. *Faraday Discuss.* **2010**, *144*, 4310.1039/B901511F
(125) Izvekov, S.; Voth, G. A Multiscale Coarse-Graining Method for Biomolecular Systems. *J. Phys. Chem. B* **2005**, *109*, 2469−2473.
(126) Cho, H.; Chu, J.-W. Inversion of radial distribution functions to pair forces by solving the Yvon−Born−Green equation iteratively. *J. Chem. Phys.* **2009**, *131*, 134107.

DOI: 10.1021/acs.chemrev.6b00163
Chem. Rev. 2016, 116, 7898−7936

(127) Lu, L.; Dama, J.; Voth, G. Fitting coarse-grained distribution functions through an iterative force-matching method. *J. Chem. Phys.* **2013**, *139*, 121906.
(128) Mullinax, J. W.; Noid, W. G. A Generalized-Yvon−Born−Green Theory for Determining Coarse-Grained Interaction Potentials†. *J. Phys. Chem. C* **2010**, *114*, 5661−5674.
(129) Shell, S. The relative entropy is fundamental to multiscale and inverse thermodynamic problems. *J. Chem. Phys.* **2008**, *129*, 144108.
(130) Bilionis, I.; Zabaras, N. A stochastic optimization approach to coarse-graining using a relative-entropy framework. *J. Chem. Phys.* **2013**, *138*, 044313.
(131) Chaimovich, A.; Shell, S. Coarse-graining errors and numerical optimization using a relative entropy framework. *J. Chem. Phys.* **2011**, *134*, 094112.
(132) Chaimovich, A.; Shell, S. Relative entropy as a universal metric for multiscale errors. *Phys. Rev. E* **2010**, *81*, 10.1103/PhysRevE.81.060104
(133) Espanol, P.; Zuniga, I. Obtaining fully dynamic coarse-grained models from MD. *Phys. Chem. Chem. Phys.* **2011**, *13*, 10538−10545.
(134) Murtola, T.; Karttunen, M.; Vattulainen, I. Systematic coarse graining from structure using internal states: application to phospholipid/cholesterol bilayer. *J. Chem. Phys.* **2009**, *131*, 05510110.1063/1.3167405
(135) Brini, E.; van der Vegt, N. F. Chemically transferable coarse-grained potentials from conditional reversible work calculations. *J. Chem. Phys.* **2012**, *137*, 15411310.1063/1.4758936
(136) Deichmann, G.; Marcon, V.; van der Vegt, N. Bottom-up derivation of conservative and dissipative interactions for coarse-grained molecular liquids with the conditional reversible work method. *J. Chem. Phys.* **2014**, *141*, 224109.
(137) Hadley, K.; McCabe, C. Coarse-Grained Molecular Models of Water: A Review. *Mol. Simul.* **2012**, *38*, 671−681.
(138) Tanaka, S.; Scheraga, H. A. Medium- and long-range interaction parameters between amino acids for predicting three-dimensional structures of proteins. *Macromolecules* **1976**, *9*, 945−950.
(139) Samudrala, R.; Moult, J. An all-atom distance-dependent conditional probability discriminatory function for protein structure prediction. *J. Mol. Biol.* **1998**, *275*, 895−916.
(140) Simons, K. T.; Ruczinski, I.; Kooperberg, C.; Fox, B. A.; Bystroff, C.; Baker, D. Improved recognition of native-like protein structures using a combination of sequence-dependent and sequence-independent features of proteins. *Proteins: Struct., Funct., Genet.* **1999**, *34*, 82−95.
(141) Xia, Y.; Levitt, M. Extracting knowledge-based energy functions from protein structures by error rate minimization: Comparison of methods using lattice model. *J. Chem. Phys.* **2000**, *113*, 9318−9330.
(142) Sippl, M. J. Boltzmann's principle, knowledge-based mean fields and protein folding. An approach to the computational determination of protein structures. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 473−501.
(143) Skolnick, J.; Kolinski, A.; Ortiz, A. Derivation of protein-specific pair potentials based on weak sequence fragment similarity. *Proteins: Struct., Funct., Genet.* **2000**, *38*, 3−16.
(144) Godzik, A.; Kolinski, A.; Skolnick, J. Are proteins ideal mixtures of amino acids? Analysis of energy parameter sets. *Protein Sci.* **1995**, *4*, 2107−2117.
(145) Skolnick, J.; Jaroszewski, L.; Kolinski, A.; Godzik, A. Derivation and testing of pair potentials for protein folding. When is the quasichemical approximation correct? *Protein Sci.* **1997**, *6*, 676−688.
(146) Majek, P.; Elber, R. A coarse-grained potential for fold recognition and molecular dynamics simulations of proteins. *Proteins: Struct., Funct., Genet.* **2009**, *76*, 822−836.
(147) Zhou, Y.; Zhou, H.; Zhang, C.; Liu, S. What is a desirable statistical energy function for proteins and how can it be obtained? *Cell Biochem. Biophys.* **2006**, *46*, 165−174.
(148) Shen, M.-Y.; Sali, A. Statistical potential for assessment and prediction of protein structures. *Protein Sci.* **2006**, *15*, 2507−2524.
(149) Yang, Y.; Zhou, Y. Specific interactions for ab initio folding of protein terminal regions with secondary structures. *Proteins: Struct., Funct., Genet.* **2008**, *72*, 793−803.
(150) Zhou, H.; Skolnick, J. GOAP: a generalized orientation-dependent, all-atom statistical potential for protein structure prediction. *Biophys. J.* **2011**, *101*, 2043−2052.
(151) Park, J.; Saitou, K. ROTAS: a rotamer-dependent, atomic statistical potential for assessment and prediction of protein structures. *BMC Bioinf.* **2014**, *15*, 307.
(152) Gront, D.; Kolinski, A. A new approach to prediction of short-range conformational propensities in proteins. *Bioinformatics* **2005**, *21*, 981−987.
(153) Amir, E.-A.; Kalisman, N.; Keasar, C. Differentiable, multi-dimensional, knowledge-based energy terms for torsion angle probabilities and propensities. *Proteins: Struct., Funct., Genet.* **2008**, *72*, 62−73.
(154) Levy-Moonshine, A.; Amir, E.-a.; Keasar, C. Enhancement of beta-sheet assembly by cooperative hydrogen bonds potential. *Bioinformatics* **2009**, *25*, 2639−2645.
(155) Liwo, A.; Arlukowicz, P.; Czaplewski, C.; Oldziej, S.; Pillardy, J.; Scheraga, H. A. A method for optimizing potential-energy functions by a hierarchical design of the potential-energy landscape: application to the UNRES force field. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 1937−1942.
(156) Mirny, L. A.; Shakhnovich, E. I. How to Derive a Protein Folding Potential? A New Approach to an Old Problem. *J. Mol. Biol.* **1996**, *264*, 1164−1179.
(157) Maiorov, V. N.; Grippen, G. M. Contact Potential That Recognizes the Correct Folding of Globular-Proteins. *J. Mol. Biol.* **1992**, *227*, 876−888.
(158) Chiu, T.-L.; Goldstein, R. A. Optimizing energy potentials for success in protein tertiary structure prediction. *Folding Des.* **1998**, *3*, 223−228.
(159) Hao, M.-H.; Scheraga, H. A. Optimizing Potential Functions for Protein Folding. *J. Phys. Chem.* **1996**, *100*, 14540−14548.
(160) Sippl, M. J. Calculation of conformational ensembles from potentials of mean force. An approach to the knowledge-based prediction of local structures in globular proteins. *J. Mol. Biol.* **1990**, *213*, 859−883.
(161) Hamelryck, T.; Borg, M.; Paluszewski, M.; Paulsen, J.; Frellsen, J.; Andreetta, C.; Boomsma, W.; Bottaro, S.; Ferkinghoff-Borg, J. Potentials of Mean Force for Protein Structure Prediction Vindicated, Formalized and Generalized. *PLoS One* **2010**, *5*, e13714.
(162) Gniewek, P.; Leelananda, S. P.; Kolinski, A.; Jernigan, R. L.; Kloczkowski, A. Multibody coarse-grained potentials for native structure recognition and quality assessment of protein models. *Proteins: Struct., Funct., Genet.* **2011**, *79*, 1923−1929.
(163) Elhefnawy, W.; Chen, L.; Han, Y.; Li, Y. ICOSA: A Distance-Dependent, Orientation-Specific Coarse-Grained Contact Potential for Protein Structure Modeling. *J. Mol. Biol.* **2015**, *427*, 2562−2576.
(164) Sankar, K.; Liu, J.; Wang, Y.; Jernigan, R. L. Distributions of experimental protein structures on coarse-grained free energy landscapes. *J. Chem. Phys.* **2015**, *143*, 243153.
(165) Zaborowski, B.; Jagieła, D.; Czaplewski, C.; Hałabis, A.; Lewandowska, A.; Żmudzińska, W.; Ołdziej, S.; Karczyńska, A.; Omieczynski, C.; Wirecki, T.; et al. A Maximum-Likelihood Approach to Force-Field Calibration. *J. Chem. Inf. Model.* **2015**, *55*, 2050−2070.
(166) Rykunov, D.; Fiser, A. New statistical potential for quality assessment of protein models and a survey of energy functions. *BMC Bioinf.* **2010**, *11*, 128.
(167) Fornes, O.; Garcia-Garcia, J.; Bonet, J.; Oliva, B. On the use of knowledge-based potentials for the evaluation of models of protein-protein, protein-DNA, and protein-RNA interactions. *Adv. Protein Chem. Struct. Biol.* **2014**, *94*, 77−120.
(168) Lu, M.; Dousis, A. D.; Ma, J. OPUS-PSP: an orientation-dependent statistical all-atom potential derived from side-chain packing. *J. Mol. Biol.* **2008**, *376*, 288−301.

DOI: 10.1021/acs.chemrev.6b00163
Chem. Rev. 2016, 116, 7898−7936

(169) Buchete, N. V.; Straub, J. E.; Thirumalai, D. Anisotropic coarse-grained statistical potentials improve the ability to identify nativelike protein structures. *J. Chem. Phys.* **2003**, *118*, 7658−7671.
(170) Zhang, Y.; Kolinski, A.; Skolnick, J. TOUCHSTONE II: a new approach to ab initio protein structure prediction. *Biophys. J.* **2003**, *85*, 1145−1164.
(171) Kolinski, A.; Bujnicki, J. M. Generalized protein structure prediction based on combination of fold-recognition with de novo folding and evaluation of models. *Proteins: Struct., Funct., Genet.* **2005**, *61*, 84−90.
(172) Xu, D.; Zhang, J.; Roy, A.; Zhang, Y. Automated protein structure modeling in CASP9 by I-TASSER pipeline combined with QUARK-based ab initio folding and FG-MD-based structure refinement. *Proteins: Struct., Funct., Genet.* **2011**, *79*, 147−160.
(173) Roy, A.; Kucukural, A.; Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* **2010**, *5*, 725−738.
(174) Taketomi, H.; Ueda, Y.; Go, N. Studies on protein folding, unfolding and fluctuations by computer simulation. I. The effect of specific amino acid sequence represented by specific inter-unit interactions. *Int. J. Pept. Protein Res.* **1975**, *7*, 445−459.
(175) Go, N.; Taketomi, H. Studies on protein folding, unfolding and fluctuations by computer simulation. IV. Hydrophobic interactions. *Int. J. Pept. Protein Res.* **1979**, *13*, 447−461.
(176) Hills, R. D.; Brooks, C. L. Insights from Coarse-Grained Go Models for Protein Folding and Dynamics. *Int. J. Mol. Sci.* **2009**, *10*, 889−905.
(177) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins: Struct., Funct., Genet.* **1995**, *21*, 167−195.
(178) Hills, R. D., Jr.; Brooks, C. L., 3rd Insights from coarse-grained Go models for protein folding and dynamics. *Int. J. Mol. Sci.* **2009**, *10*, 889−905.
(179) Lammert, H.; Schug, A.; Onuchic, J. N. Robustness and generalization of structure-based models for protein folding and function. *Proteins: Struct., Funct., Genet.* **2009**, *77*, 881−891.
(180) Hills, R. D., Jr. Balancing bond, nonbond, and go-like terms in coarse grain simulations of conformational dynamics. *Methods Mol. Biol.* **2014**, *1084*, 123−140.
(181) Christen, M.; van Gunsteren, W. On searching in, sampling of, and dynamically moving through conformational space of biomolecular systems: A review. *J. Comput. Chem.* **2008**, *29*, 157−166.
(182) Liwo, A.; Czaplewski, C.; Oldziej, S.; Scheraga, H. A. Computational techniques for efficient conformational sampling of proteins. *Curr. Opin. Struct. Biol.* **2008**, *18*, 134−139.
(183) Spiwok, V.; Sucur, Z.; Hosek, P. Enhanced sampling techniques in biomolecular simulations. *Biotechnol. Adv.* **2015**, *33*, 1130−1140.
(184) Bernardi, R. C.; Melo, M. C.; Schulten, K. Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochim. Biophys. Acta, Gen. Subj.* **2015**, *1850*, 872−877.
(185) Hatfield, M. P.; Lovas, S. Conformational sampling techniques. *Curr. Pharm. Des.* **2014**, *20*, 3303−3313.
(186) Lee, J.; Scheraga, H. A.; Rackovsky, S. New optimization method for conformational energy calculations on polypeptides: Conformational space annealing. *J. Comput. Chem.* **1997**, *18*, 1222−1232.
(187) Stein, A.; Kortemme, T. Improvements to robotics-inspired conformational sampling in rosetta. *PLoS One* **2013**, *8*, e63090.
(188) Shmygelska, A.; Hoos, H. H. An ant colony optimization algorithm for the 2D and 3D hydrophobic polar protein folding problem. *BMC Bioinf.* **2005**, *6*, 30.
(189) Peter, E. K.; Lykov, K.; Pivkin, I. V. A polarizable coarse-grained protein model for dissipative particle dynamics. *Phys. Chem. Chem. Phys.* **2015**, *17*, 24452−24461.
(190) Ermak, D.; McCammon, J. A. Brownian dynamics with hydrodynamic interactions. *J. Chem. Phys.* **1978**, *69*, 1352−1360.
(191) Proctor, E.; Dokholyan, N. Applications of Discrete Molecular Dynamics in biology and medicine. *Curr. Opin. Struct. Biol.* **2016**, *37*, 9−13.
(192) Urbanc, B.; Borreguero, J.; Cruz, L.; Stanley, E. Ab initio discrete molecular dynamics approach to protein folding and aggregation. *Methods Enzymol.* **2006**, *412*, 314−338.
(193) Chen, Y.; Ding, F.; Dokholyan, N. V. Fidelity of the protein structure reconstruction from inter-residue proximity constraints. *J. Phys. Chem. B* **2007**, *111*, 7432−7438.
(194) Sfriso, P.; Hospital, A.; Emperador, A.; Orozco, M. Exploration of conformational transition pathways from coarse-grained simulations. *Bioinformatics* **2013**, *29*, 1980−1986.
(195) Urbanc, B.; Cruz, L.; Ding, F.; Sammond, D.; Khare, S.; Buldyrev, S. V.; Stanley, H. E.; Dokholyan, N. V. Molecular dynamics simulation of amyloid beta dimer formation. *Biophys. J.* **2004**, *87*, 2310−2321.
(196) Vitalis, A.; Pappu, R. Methods for Monte Carlo simulations of biomacromolecules. *Annu. Rep. Comput. Chem.* **2009**, *5*, 49−76.
(197) Goujon, F.; Malfreyt, P.; Simon, J.-M.; Boutin, A.; Rousseau, B.; Fuchs, A. Monte Carlo versus molecular dynamics simulations in heterogeneous systems: An application to the n-pentane liquid-vapor interface. *J. Chem. Phys.* **2004**, *121*, 12559−12571.
(198) Schluttig, J.; Bachmann, M.; Janke, W. Comparative molecular dynamics and Monte Carlo study of statistical properties for coarse-grained heteropolymers. *J. Comput. Chem.* **2008**, *29*, 2603−2612.
(199) Lane, T. J.; Shukla, D.; Beauchamp, K. A.; Pande, V. S. To milliseconds and beyond: challenges in the simulation of protein folding. *Curr. Opin. Struct. Biol.* **2013**, *23*, 58−65.
(200) Bowman, G. R.; Voelz, V. A.; Pande, V. S. Taming the complexity of protein folding. *Curr. Opin. Struct. Biol.* **2011**, *21*, 4−11.
(201) Noe, F.; Fischer, S. Transition networks for modeling the kinetics of conformational change in macromolecules. *Curr. Opin. Struct. Biol.* **2008**, *18*, 154−162.
(202) Yin, Y.; Sieradzan, A. K.; Liwo, A.; He, Y.; Scheraga, H. A. Physics-based potentials for coarse-grained modeling of protein-DNA interactions. *J. Chem. Theory Comput.* **2015**, *11*, 1792−1808.
(203) Krupa, P.; Mozolewska, M. A.; Joo, K.; Lee, J.; Czaplewski, C.; Liwo, A. Prediction of Protein Structure by Template-Based Modeling Combined with the UNRES Force Field. *J. Chem. Inf. Model.* **2015**, *55*, 1271−1281.
(204) Mozolewska, M. A.; Krupa, P.; Scheraga, H. A.; Liwo, A. Molecular modeling of the binding modes of the iron-sulfur protein to the Jac1 co-chaperone from Saccharomyces cerevisiae by all-atom and coarse-grained approaches. *Proteins: Struct., Funct., Genet.* **2015**, *83*, 1414−1426.
(205) Blaszczyk, M.; Jamroz, M.; Kmiecik, S.; Kolinski, A. CABS-fold: Server for the de novo and consensus-based prediction of protein structure. *Nucleic Acids Res.* **2013**, *41*, W406−411.
(206) Jamroz, M.; Kolinski, A. Modeling of loops in proteins: a multi-method approach. *BMC Struct. Biol.* **2010**, *10*, 5.
(207) Kmiecik, S.; Kolinski, A. Characterization of protein-folding pathways by reduced-space modeling. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 12330−12335.
(208) Kmiecik, S.; Kolinski, A. Folding pathway of the b1 domain of protein G explored by multiscale modeling. *Biophys. J.* **2008**, *94*, 726−736.
(209) Kmiecik, S.; Kolinski, A. Simulation of chaperonin effect on protein folding: a shift from nucleation-condensation to framework mechanism. *J. Am. Chem. Soc.* **2011**, *133*, 10283−10289.
(210) Jamroz, M.; Kolinski, A.; Kmiecik, S. CABS-flex: Server for fast simulation of protein structure fluctuations. *Nucleic Acids Res.* **2013**, *41*, W427−431.
(211) Jamroz, M.; Orozco, M.; Kolinski, A.; Kmiecik, S. Consistent View of Protein Fluctuations from All-Atom Molecular Dynamics and Coarse-Grained Dynamics with Knowledge-Based Force-Field. *J. Chem. Theory Comput.* **2013**, *9*, 119−125.
(212) Kurcinski, M.; Kolinski, A.; Kmiecik, S. Mechanism of Folding and Binding of an Intrinsically Disordered Protein As Revealed by ab Initio Simulations. *J. Chem. Theory Comput.* **2014**, *10*, 2224−2231.

DOI: 10.1021/acs.chemrev.6b00163
Chem. Rev. 2016, 116, 7898−7936

(213) Kurcinski, M.; Jamroz, M.; Blaszczyk, M.; Kolinski, A.; Kmiecik, S. CABS-dock web server for the flexible docking of peptides to proteins without prior knowledge of the binding site. *Nucleic Acids Res.* **2015**, *43*, W419−424.
(214) Blaszczyk, M.; Kurcinski, M.; Kouza, M.; Wieteska, L.; Debinski, A.; Kolinski, A.; Kmiecik, S. Modeling of protein-peptide interactions using the CABS-dock web server for binding site search and flexible docking. *Methods* **2016**, *93*, 72−83.
(215) Kar, P.; Gopal, S. M.; 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**, *10*, 3459−3472.
(216) Marrink, S. J.; de Vries, A. H.; Mark, A. E. Coarse grained model for semiquantitative lipid simulations. *J. Phys. Chem. B* **2004**, *108*, 750−760.
(217) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S. J. The MARTINI coarse-grained force field: Extension to proteins. *J. Chem. Theory Comput.* **2008**, *4*, 819−834.
(218) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. The MARTINI force field: Coarse grained model for biomolecular simulations. *J. Phys. Chem. B* **2007**, *111*, 7812−7824.
(219) Shih, A. Y.; Arkhipov, A.; Freddolino, P. L.; Schulten, K. Coarse grained protein-lipid model with application to lipoprotein particles. *J. Phys. Chem. B* **2006**, *110*, 3674−3684.
(220) Bond, P. J.; Holyoake, J.; Ivetac, A.; Khalid, S.; Sansom, M. S. Coarse-grained molecular dynamics simulations of membrane proteins and peptides. *J. Struct. Biol.* **2007**, *157*, 593−605.
(221) Spiga, E.; Alemani, D.; Degiacomi, M. T.; Cascella, M.; Peraro, M. D. Electrostatic-Consistent Coarse-Grained Potentials for Molecular Simulations of Proteins. *J. Chem. Theory Comput.* **2013**, *9*, 3515−3526.
(222) Alemani, D.; Collu, F.; Cascella, M.; Dal Peraro, M. A Nonradial Coarse-Grained Potential for Proteins Produces Naturally Stable Secondary Structure Elements. *J. Chem. Theory Comput.* **2010**, *6*, 315−324.
(223) Arnarez, C.; Uusitalo, J. J.; Masman, M. F.; Ingolfsson, H. I.; de Jong, D. H.; Melo, M. N.; Periole, X.; de Vries, A. H.; Marrink, S. J. Dry Martini, a Coarse-Grained Force Field for Lipid Membrane Simulations with Implicit Solvent. *J. Chem. Theory Comput.* **2015**, *11*, 260−275.
(224) Yesylevskyy, S. O.; Schafer, L. V.; Sengupta, D.; Marrink, S. J. Polarizable water model for the coarse-grained MARTINI force field. *PLoS Comput. Biol.* **2010**, *6*, e1000810.
(225) Marrink, S. J.; Tieleman, D. P. Perspective on the Martini model. *Chem. Soc. Rev.* **2013**, *42*, 6801−6822.
(226) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. GROMACS: fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701−1718.
(227) Bowers, K. J.; Chow, E.; Xu, H.; Dror, R. O.; Eastwood, M. P.; Gregersen, B. A.; Klepeis, J. L.; Kolossváry, I.; Moraes, M. A.; Sacerdoti, F. D. et al. *Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. Proceedings of the ACM/IEEE Conference on Supercomputing (SC06)*, 2006.
(228) Baron, R.; Trzesniak, D.; de Vries, A. H.; Elsener, A.; Marrink, S. J.; van Gunsteren, W. F. Comparison of thermodynamic properties of coarse-grained and atomic-level simulation models. *ChemPhysChem* **2007**, *8*, 452−461.
(229) Leaver-Fay, A.; Tyka, M.; Lewis, S.; Lange, O.; Thompson, J.; Jacak, R.; Kaufman, K.; Renfrew, D.; Smith, C.; Sheffler, W.; et al. ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* **2011**, *487*, 545−574.
(230) Bradley, P.; Misura, K. M.; Baker, D. Toward high-resolution de novo structure prediction for small proteins. *Science* **2005**, *309*, 1868−1871.
(231) Gront, D.; Kulp, D.; Vernon, R.; Strauss, C.; Baker, D. Generalized fragment picking in rosetta: design, protocols and applications. *PLoS One* **2011**, *6*, e23294.
(232) Smith, C.; Kortemme, T. Backrub-Like Backbone Simulation Recapitulates Natural Protein Conformational Variability and Improves Mutant Side-Chain Prediction. *J. Mol. Biol.* **2008**, *380*, 742−756.
(233) Betancourt, M. Efficient Monte Carlo trial moves for polypeptide simulations. *J. Chem. Phys.* **2005**, *123*, 174905.
(234) Dunbrack, Jr.; Karplus, M. Backbone-dependent Rotamer Library for Proteins Application to Side-chain Prediction. *J. Mol. Biol.* **1993**, *230*, 543−574.
(235) Kuhlman, B.; Dantas, G.; Ireton, G.; Varani, G.; Stoddard, B.; Baker, D. Design of a novel globular protein fold with atomic-level accuracy. *Science* **2003**, *302*, 1364−1368.
(236) Davtyan, A.; Schafer, N. P.; Zheng, W.; Clementi, C.; Wolynes, P. G.; Papoian, G. A. AWSEM-MD: protein structure prediction using coarse-grained physical potentials and bioinformatically based local structure biasing. *J. Phys. Chem. B* **2012**, *116*, 8494−8503.
(237) Zheng, W.; Schafer, N. P.; Davtyan, A.; Papoian, G. A.; Wolynes, P. G. Predictive energy landscapes for protein-protein association. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 19244−19249.
(238) Zheng, W.; Schafer, N. P.; Wolynes, P. G. Frustration in the energy landscapes of multidomain protein misfolding. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 1680−1685.
(239) Chen, M.; Zheng, W.; Wolynes, P. G. Energy landscapes of a mechanical prion and their implications for the molecular mechanism of long-term memory. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 5006.
(240) Tsai, M. Y.; Zheng, W.; Balamurugan, D.; Schafer, N. P.; Kim, B. L.; Cheung, M. S.; Wolynes, P. G. Electrostatics, structure prediction, and the energy landscapes for protein folding and binding. *Protein Sci.* **2016**, *25*, 255−269.
(241) Kim, B. L.; Schafer, N. P.; Wolynes, P. G. Predictive energy landscapes for folding alpha-helical transmembrane proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 11031−11036.
(242) Bereau, T.; Deserno, M. Generic coarse-grained model for protein folding and aggregation. *J. Chem. Phys.* **2009**, *130*, 235106.
(243) 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.
(244) Arnold, A.; Lenz, O.; Kesselheim, S.; Weeber, R.; Fahrenberger, F.; Roehm, D.; Košovan, P.; Holm, C. In *Meshfree Methods for Partial Differential Equations VI*; Griebel, M., Schweitzer, M. A., Eds.; Springer: Berlin, 2013; Vol. *89*.
(245) Pronk, S.; Pall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; et al. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29*, 845−854.
(246) Sterpone, F.; Melchionna, S.; Tuffery, P.; Pasquali, S.; Mousseau, N.; Cragnolini, T.; Chebaro, Y.; St-Pierre, J. F.; Kalimeri, M.; Barducci, A.; et al. The OPEP protein model: from single molecules, amyloid formation, crowding and hydrodynamics to DNA/RNA systems. *Chem. Soc. Rev.* **2014**, *43*, 4871−4893.
(247) Sterpone, F.; Derreumaux, P.; Melchionna, S. Protein Simulations in Fluids: Coupling the OPEP Coarse-Grained Force Field with Hydrodynamics. *J. Chem. Theory Comput.* **2015**, *11*, 1843−1853.
(248) Barducci, A.; Bonomi, M.; Derreumaux, P. Assessing the Quality of the OPEP Coarse-Grained Force Field. *J. Chem. Theory Comput.* **2011**, *7*, 1928−1934.
(249) 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.
(250) Chebaro, Y.; Dong, X.; Laghaei, R.; Derreumaux, P.; Mousseau, N. Replica exchange molecular dynamics simulations of coarse-grained proteins in implicit solvent. *J. Phys. Chem. B* **2009**, *113*, 267−274.
(251) Nasica-Labouze, J.; Nguyen, P. H.; Sterpone, F.; Berthoumieu, O.; Buchete, N. V.; Cote, S.; De Simone, A.; Doig, A. J.; Faller, P.; Garcia, A.; et al. Amyloid beta Protein and Alzheimer's Disease: When Computer Simulations Complement Experimental Studies. *Chem. Rev.* **2015**, *115*, 3518−3563.

DOI: 10.1021/acs.chemrev.6b00163
Chem. Rev. 2016, 116, 7898−7936

(252) Nasica-Labouze, J.; Meli, M.; Derreumaux, P.; Colombo, G.; Mousseau, N. A multiscale approach to characterize the early aggregation steps of the amyloid-forming peptide GNNQQNY from the yeast prion sup-35. *PLoS Comput. Biol.* **2011**, *7*, e1002051.
(253) Maupetit, J.; Derreumaux, P.; Tuffery, P. A fast method for large-scale de novo peptide and miniprotein structure prediction. *J. Comput. Chem.* **2010**, *31*, 726−738.
(254) Maupetit, J.; Derreumaux, P.; Tuffery, P. PEP-FOLD: an online resource for de novo peptide structure prediction. *Nucleic Acids Res.* **2009**, *37*, W498−503.
(255) Pasi, M.; Lavery, R.; Ceres, N. PaLaCe: A Coarse-Grain Protein Model for Studying Mechanical Properties. *J. Chem. Theory Comput.* **2013**, *9*, 785−793.
(256) Frezza, E.; Lavery, R. Internal Normal Mode Analysis (iNMA) Applied to Protein Conformational Flexibility. *J. Chem. Theory Comput.* **2015**, *11*, 5503−5512.
(257) Hinsen, K. The molecular modeling toolkit: A new approach to molecular simulations. *J. Comput. Chem.* **2000**, *21*, 79−85.
(258) Feig, M.; Karanicolas, J.; Brooks, C. L., 3rd MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. *J. Mol. Graphics Modell.* **2004**, *22*, 377−395.
(259) Fleishman, S. J.; Corn, J. E.; Strauch, E. M.; Whitehead, T. A.; Andre, I.; Thompson, J.; Havranek, J. J.; Das, R.; Bradley, P.; Baker, D. Rosetta in CAPRI rounds 13−19. *Proteins: Struct., Funct., Genet.* **2010**, *78*, 3212−3218.
(260) Davis, I. W.; Baker, D. RosettaLigand docking with full ligand and receptor flexibility. *J. Mol. Biol.* **2009**, *385*, 381−392.
(261) Weitzner, B. D.; Kuroda, D.; Marze, N.; Xu, J.; Gray, J. J. Blind prediction performance of RosettaAntibody 3.0: grafting, relaxation, kinematic loop modeling, and full CDR optimization. *Proteins: Struct., Funct., Genet.* **2014**, *82*, 1611−1623.
(262) DiMaio, F.; Echols, N.; Headd, J. J.; Terwilliger, T. C.; Adams, P. D.; Baker, D. Improved low-resolution crystallographic refinement with Phenix and Rosetta. *Nat. Methods* **2013**, *10*, 1102−1104.
(263) Mao, B. C.; Tejero, R.; Baker, D.; Montelione, G. T. Protein NMR Structures Refined with Rosetta Have Higher Accuracy Relative to Corresponding X-ray Crystal Structures. *J. Am. Chem. Soc.* **2014**, *136*, 1893−1906.
(264) London, N.; Raveh, B.; Cohen, E.; Fathi, G.; Schueler-Furman, O. Rosetta FlexPepDock web server–high resolution modeling of peptide-protein interactions. *Nucleic Acids Res.* **2011**, *39*, W249−253.
(265) Liu, T.; Pan, X.; Chao, L.; Tan, W.; Qu, S.; Yang, L.; Wang, B.; Mei, H. Subangstrom accuracy in pHLA-I modeling by Rosetta FlexPepDock refinement protocol. *J. Chem. Inf. Model.* **2014**, *54*, 2233−2242.
(266) Thyme, S.; Baker, D. Redesigning the specificity of protein-DNA interactions with Rosetta. *Methods Mol. Biol.* **2014**, *1123*, 265−282.
(267) Lyskov, S.; Chou, F. C.; Conchuir, S. O.; Der, B. S.; Drew, K.; Kuroda, D.; Xu, J.; Weitzner, B. D.; Renfrew, P. D.; Sripakdeevong, P.; et al. Serverification of molecular modeling applications: the Rosetta Online Server that Includes Everyone (ROSIE). *PLoS One* **2013**, *8*, e63906.
(268) 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.
(269) Liwo, A.; Khalili, M.; Scheraga, H. A. Ab initio simulations of protein-folding pathways by molecular dynamics with the united-residue model of polypeptide chains. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 2362−2367.
(270) Zhou, R.; Maisuradze, G. G.; Sunol, D.; Todorovski, T.; Macias, M. J.; Xiao, Y.; Scheraga, H. A.; Czaplewski, C.; Liwo, A. Folding kinetics of WW domains with the united residue force field for bridging microscopic motions and experimental measurements. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 18243−18248.
(271) Maisuradze, G. G.; Senet, P.; Czaplewski, C.; Liwo, A.; Scheraga, H. A. Investigation of protein folding by coarse-grained molecular dynamics with the UNRES force field. *J. Phys. Chem. A* **2010**, *114*, 4471−4485.
(272) Kachlishvili, K.; Maisuradze, G. G.; Martin, O. A.; Liwo, A.; Vila, J. A.; Scheraga, H. A. Accounting for a mirror-image conformation as a subtle effect in protein folding. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 8458−8463.
(273) He, Y.; Mozolewska, M. A.; Krupa, P.; Sieradzan, A. K.; Wirecki, T. K.; Liwo, A.; Kachlishvili, K.; Rackovsky, S.; Jagiela, D.; Slusarz, R.; et al. Lessons from application of the UNRES force field to predictions of structures of CASP10 targets. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 14936−14941.
(274) Krupa, P.; Sieradzan, A. K.; Rackovsky, S.; Baranowski, M.; Oldziej, S.; Scheraga, H. A.; Liwo, A.; Czaplewski, C. Improvement of the treatment of loop structures in the UNRES force field by inclusion of coupling between backbone- and side-chain-local conformational states. *J. Chem. Theory Comput.* **2013**, *9*, 462010.1021/ct4004977
(275) Sieradzan, A. K.; Liwo, A.; Hansmann, U. H. Folding and self-assembly of a small protein complex. *J. Chem. Theory Comput.* **2012**, *8*, 3416−3422.
(276) Rojas, A.; Liwo, A.; Browne, D.; Scheraga, H. A. Mechanism of fiber assembly: treatment of Abeta peptide aggregation with a coarse-grained united-residue force field. *J. Mol. Biol.* **2010**, *404*, 537−552.
(277) Golas, E.; Maisuradze, G. G.; Senet, P.; Oldziej, S.; Czaplewski, C.; Scheraga, H. A.; Liwo, A. Simulation of the opening and closing of Hsp70 chaperones by coarse-grained molecular dynamics. *J. Chem. Theory Comput.* **2012**, *8*, 1750−1764.
(278) Boras, B. W.; Hirakis, S. P.; Votapka, L. W.; Malmstrom, R. D.; Amaro, R. E.; McCulloch, A. D. Bridging scales through multiscale modeling: a case study on protein kinase A. *Front. Physiol.* **2015**, *6*, 250.
(279) Ackerman, D. G.; Feigenson, G. W. Multiscale modeling of four-component lipid mixtures: domain composition, size, alignment, and properties of the phase interface. *J. Phys. Chem. B* **2015**, *119*, 4240−4250.
(280) Hills, R. D., Jr.; McGlinchey, N. Model parameters for simulation of physiological lipids. *J. Comput. Chem.* **2016**, *37*, 1112.
(281) Bromfield, S. M.; Posocco, P.; Fermeglia, M.; Tolosa, J.; Herreros-Lopez, A.; Pricl, S.; Rodriguez-Lopez, J.; Smith, D. K. Shape-persistent and adaptive multivalency: rigid transgeden (TGD) and flexible PAMAM dendrimers for heparin binding. *Chem. - Eur. J.* **2014**, *20*, 9666−9674.
(282) Zavadlav, J.; Podgornik, R.; Praprotnik, M. Adaptive Resolution Simulation of a DNA Molecule in Salt Solution. *J. Chem. Theory Comput.* **2015**, *11*, 5035−5044.
(283) Panteva, M. T.; Dissanayake, T.; Chen, H.; Radak, B. K.; Kuechler, E. R.; Giambasu, G. M.; Lee, T. S.; York, D. M. Multiscale methods for computational RNA enzymology. *Methods Enzymol.* **2015**, *553*, 335−374.
(284) Ayton, G. S.; Voth, G. A. Systematic multiscale simulation of membrane protein systems. *Curr. Opin. Struct. Biol.* **2009**, *19*, 138−144.
(285) Sharma, S.; Ding, F.; Dokholyan, N. V. Multiscale modeling of nucleosome dynamics. *Biophys. J.* **2007**, *92*, 1457−1470.
(286) Gao, T.; Blackwell, R.; Glaser, M. A.; Betterton, M. D.; Shelley, M. J. Multiscale modeling and simulation of microtubule-motor-protein assemblies. *Phys. Rev. E Stat. Nonlin. Soft. Matter Phys.* **2015**, *92*, 062709.
(287) Xie, Z. R.; Chen, J.; Wu, Y. Multiscale Model for the Assembly Kinetics of Protein Complexes. *J. Phys. Chem. B* **2016**, *120*, 621.
(288) Zheng, J. J.; Wang, D. Q.; Thiel, W.; Shaik, S. QM/MM study of mechanisms for compound I formation in the catalytic cycle of cytochrome P450cam. *J. Am. Chem. Soc.* **2006**, *128*, 13204−13215.
(289) Rothlisberger, U.; Carloni, P.; Doclo, K.; Parrinello, M. A comparative study of galactose oxidase and active site analogs based on QM/MM Car Parrinello simulations. *JBIC, J. Biol. Inorg. Chem.* **2000**, *5*, 236−250.
(290) Shurki, A.; Warshel, A. Structure/function correlations of proteins using MM, QM/MM, and related approaches: Methods, concepts, pitfalls, and current progress. *Adv. Protein Chem.* **2003**, *66*, 249−313.

DOI: 10.1021/acs.chemrev.6b00163
Chem. Rev. 2016, 116, 7898−7936

(291) Roy, S.; Kastner, J. Synergistic Substrate and Oxygen Activation in Salicylate Dioxygenase Revealed by QM/MM Simulations. *Angew. Chem., Int. Ed.* **2016**, *55*, 1168−1172.
(292) Raich, L.; Borodkin, V.; Fang, W.; Castro-Lopez, J.; van Aalten, D. M.; Hurtado-Guerrero, R.; Rovira, C. A trapped covalent intermediate of a glycoside hydrolase on the pathway to trans-glycosylation. Insights from experiments and QM/MM simulations. *J. Am. Chem. Soc.* **2016**, *138*, 3325.
(293) Kumar, R. P.; Roopa, L.; Nongthomba, U.; Sudheer Mohammed, M. M.; Kulkarni, N. Docking, molecular dynamics and QM/MM studies to delineate the mode of binding of CucurbitacinE to F-actin. *J. Mol. Graphics Modell.* **2016**, *63*, 29−37.
(294) Carvalho, A. T.; Barrozo, A.; Doron, D.; Kilshtain, A. V.; Major, D. T.; Kamerlin, S. C. Challenges in computational studies of enzyme structure, function and dynamics. *J. Mol. Graphics Modell.* **2014**, *54*, 62−79.
(295) Duarte, F.; Amrein, B. A.; Blaha-Nelson, D.; Kamerlin, S. C. Recent advances in QM/MM free energy calculations using reference potentials. *Biochim. Biophys. Acta, Gen. Subj.* **2015**, *1850*, 954−965.
(296) Senn, H. M.; Thiel, W. QM/MM methods for biomolecular systems. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198−1229.
(297) Quesne, M. G.; Borowski, T.; de Visser, S. P. Quantum Mechanics/Molecular Mechanics Modeling of Enzymatic Processes: Caveats and Breakthroughs. *Chem. - Eur. J.* **2016**, *22*, 2562.
(298) Kamerlin, S. C.; Haranczyk, M.; Warshel, A. Progress in ab initio QM/MM free-energy simulations of electrostatic energies in proteins: accelerated QM/MM studies of pKa, redox reactions and solvation free energies. *J. Phys. Chem. B* **2009**, *113*, 1253−1272.
(299) Heath, A. P.; Kavraki, L. E.; Clementi, C. From coarse-grain to all-atom: toward multiscale analysis of protein landscapes. *Proteins: Struct., Funct., Genet.* **2007**, *68*, 646−661.
(300) Machado, M. R.; Dans, P. D.; Pantano, S. A hybrid all-atom/coarse grain model for multiscale simulations of DNA. *Phys. Chem. Chem. Phys.* **2011**, *13*, 18134−18144.
(301) Michel, J.; Orsi, M.; Essex, J. W. Prediction of partition coefficients by multiscale hybrid atomic-level/coarse-grain simulations. *J. Phys. Chem. B* **2008**, *112*, 657−660.
(302) Vieth, M.; Kolinski, A.; Brooks, C. L., 3rd; Skolnick, J. Prediction of the folding pathways and structure of the GCN4 leucine zipper. *J. Mol. Biol.* **1994**, *237*, 361−367.
(303) Mohanty, D.; Dominy, B. N.; Kolinski, A.; Brooks, C. L., 3rd; Skolnick, J. Correlation between knowledge-based and detailed atomic potentials: application to the unfolding of the GCN4 leucine zipper. *Proteins: Struct., Funct., Genet.* **1999**, *35*, 447−452.
(304) Fan, Z. Z.; Hwang, J. K.; Warshel, A. Using simplified protein representation as a reference potential for all-atom calculations of folding free energy. *Theor. Chem. Acc.* **1999**, *103*, 77−80.
(305) Zhang, J.; Li, W.; Wang, J.; Qin, M.; Wu, L.; Yan, Z.; Xu, W.; Zuo, G.; Wang, W. Protein folding simulations: from coarse-grained model to all-atom model. *IUBMB Life* **2009**, *61*, 627−643.
(306) Reith, D.; Putz, M.; Muller-Plathe, F. Deriving effective mesoscale potentials from atomistic simulations. *J. Comput. Chem.* **2003**, *24*, 1624−1636.
(307) Chu, J. W.; Voth, G. A. Coarse-grained modeling of the actin filament derived from atomistic-scale simulations. *Biophys. J.* **2006**, *90*, 1572−1582.
(308) Shi, Q.; Izvekov, S.; Voth, G. A. Mixed atomistic and coarse-grained molecular dynamics: simulation of a membrane-bound ion channel. *J. Phys. Chem. B* **2006**, *110*, 15045−15048.
(309) Lyman, E.; Zuckerman, D. M. Resolution Exchange Simulation with Incremental Coarsening. *J. Chem. Theory Comput.* **2006**, *2*, 656−666.
(310) Villa, E.; Balaeff, A.; Schulten, K. Structural dynamics of the lac repressor-DNA complex revealed by a multiscale simulation. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6783−6788.
(311) 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*, 05411710.1063/1.4817209
(312) Orsi, M.; Sanderson, W. E.; Essex, J. W. Permeability of Small Molecules through a Lipid Bilayer: A Multiscale Simulation Study. *J. Phys. Chem. B* **2009**, *113*, 12019−12029.
(313) Neri, M.; Anselmi, C.; Carnevale, V.; Vargiu, A. V.; Carloni, P. Molecular dynamics simulations of outer-membrane protease T from E-coli based on a hybrid coarse-grained/atomistic potential. *J. Phys.: Condens. Matter* **2006**, *18*, S347−S355.
(314) 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.
(315) 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*, 110.
(316) Heyden, A.; Lin, H.; Truhlar, D. G. Adaptive partitioning in combined quantum mechanical and molecular mechanical calculations of potential energy functions for multiscale simulations. *J. Phys. Chem. B* **2007**, *111*, 2231−2241.
(317) Kerdcharoen, T.; Liedl, K. R.; Rode, B. M. A QM/MM simulation method applied to the solution of Li+ in liquid ammonia. *Chem. Phys.* **1996**, *211*, 313−323.
(318) Hofer, T. S.; Pribil, A. B.; Randolf, B. R.; Rode, B. M. Structure and dynamics of solvated Sn(II) in aqueous solution: An ab initio QM/MM MD approach. *J. Am. Chem. Soc.* **2005**, *127*, 14231−14238.
(319) Kerdcharoen, T.; Morokuma, K. ONIOM-XS: an extension of the ONIOM method for molecular simulation in condensed phase. *Chem. Phys. Lett.* **2002**, *355*, 257−262.
(320) Iwata, Y.; Kasuya, A.; Miyamoto, S. An efficient method for reconstructing protein backbones from [alpha]-carbon coordinates. *J. Mol. Graphics Modell.* **2002**, *21*, 119−128.
(321) Purisima, E. O.; Scheraga, H. A. Conversion from a virtual-bond chain to a complete polypeptide backbone chain. *Biopolymers* **1984**, *23*, 1207−1224.
(322) Gront, D.; Kmiecik, S.; Kolinski, A. Backbone building from quadrilaterals: a fast and accurate algorithm for protein backbone reconstruction from alpha carbon coordinates. *J. Comput. Chem.* **2007**, *28*, 1593−1597.
(323) Rotkiewicz, P.; Skolnick, J. Fast procedure for reconstruction of full-atom protein models from reduced representations. *J. Comput. Chem.* **2008**, *29*, 1460−1465.
(324) Feig, M.; Rotkiewicz, P.; Kolinski, A.; Skolnick, J.; Brooks, C. L., 3rd Accurate reconstruction of all-atom protein representations from side-chain-based low-resolution models. *Proteins: Struct., Funct., Genet.* **2000**, *41*, 86−97.
(325) Milik, M.; Kolinski, A.; Skolnick, J. Algorithm for rapid reconstruction of protein backbone from alpha carbon coordinates. *J. Comput. Chem.* **1997**, *18*, 80−85.
(326) Adcock, S. Peptide backbone reconstruction using dead-end elimination and a knowledge-based forcefield. *J. Comput. Chem.* **2004**, *25*, 16−27.
(327) Claessens, M.; van Cutsem, E.; Lasters, I.; Wodak, S. Modelling the polypeptide backbone with 'spare parts' from known protein structures. *Protein Eng., Des. Sel.* **1989**, *2*, 335−345.
(328) Pierce, N. A.; Winfree, E. Protein design is NP-hard. *Protein Eng., Des. Sel.* **2002**, *15*, 779−782.
(329) Gordon, D. B.; Hom, G. K.; Mayo, S. L.; Pierce, N. A. Exact rotamer optimization for protein design. *J. Comput. Chem.* **2003**, *24*, 232−243.
(330) Bower, M. J.; Cohen, F. E.; Dunbrack, R. L., Jr. Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: a new homology modeling tool. *J. Mol. Biol.* **1997**, *267*, 1268−1282.
(331) Subramaniam, S.; Senes, A. Backbone dependency further improves side chain prediction efficiency in the Energy-based Conformer Library (bEBL). *Proteins: Struct., Funct., Genet.* **2014**, *82*, 3177−3187.
(332) Holm, L.; Sander, C. Fast and simple Monte Carlo algorithm for side chain optimization in proteins: application to model building by homology. *Proteins: Struct., Funct., Genet.* **1992**, *14*, 213−223.

DOI: 10.1021/acs.chemrev.6b00163
Chem. Rev. 2016, 116, 7898−7936

(333) Liang, S.; Grishin, N. V. Side-chain modeling with an optimized scoring function. *Protein Sci.* **2002**, *11*, 322−331.
(334) Desmet, J.; De Maeyer, M.; Hazes, B.; Lasters, I. The dead-end elimination theorem and its use in protein side-chain positioning. *Nature* **1992**, *356*, 539−542.
(335) Looger, L. L.; Hellinga, H. W. Generalized dead-end elimination algorithms make large-scale protein side-chain structure prediction tractable: implications for protein design and structural genomics. *J. Mol. Biol.* **2001**, *307*, 429−445.
(336) de Armas, H. N.; Dewilde, M.; Verbeke, K.; De Maeyer, M.; Declerck, P. J. Study of recombinant antibody fragments and PAI-1 complexes combining protein-protein docking and results from site-directed mutagenesis (vol 15, pg 1105, 2007). *Structure* **2007**, *15*, 1339−1339.
(337) Lee, C.; Subbiah, S. Prediction of protein side-chain conformation by packing optimization. *J. Mol. Biol.* **1991**, *217*, 373−388.
(338) Ravindranath, P. A.; Forli, S.; Goodsell, D. S.; Olson, A. J.; Sanner, M. F. AutoDockFR: Advances in Protein-Ligand Docking with Explicitly Specified Binding Site Flexibility. *PLoS Comput. Biol.* **2015**, *11*, e1004586.
(339) Pedersen, J. T.; Moult, J. Genetic algorithms for protein structure prediction. *Curr. Opin. Struct. Biol.* **1996**, *6*, 227−231.
(340) Kingsford, C. L.; Chazelle, B.; Singh, M. Solving and analyzing side-chain positioning problems using linear and integer programming. *Bioinformatics* **2005**, *21*, 1028−1036.
(341) Xu, J.; Jiao, F.; Berger, B. A tree-decomposition approach to protein structure prediction. *Proc. IEEE Comput. Syst. Bioinform. Conf.* **2005**, 247−256.
(342) Canutescu, A. A.; Shelenkov, A. A.; Dunbrack, R. L., Jr. A graph-theory algorithm for rapid protein side-chain prediction. *Protein Sci.* **2003**, *12*, 2001−2014.
(343) Nagata, K.; Randall, A.; Baldi, P. SIDEpro: a novel machine learning approach for the fast and accurate prediction of side-chain conformations. *Proteins: Struct., Funct., Genet.* **2012**, *80*, 142−153.
(344) Mendes, J.; Baptista, A. M.; Carrondo, M. A.; Soares, C. M. Improved modeling of side-chains in proteins with rotamer-based methods: a flexible rotamer model. *Proteins: Struct., Funct., Genet.* **1999**, *37*, 530−543.
(345) Koehl, P.; Delarue, M. Application of a self-consistent mean field theory to predict protein side-chains conformation and estimate their conformational entropy. *J. Mol. Biol.* **1994**, *239*, 249−275.
(346) Desmet, J.; Spriet, J.; Lasters, I. Fast and accurate side-chain topology and energy refinement (FASTER) as a new method for protein structure optimization. *Proteins: Struct., Funct., Genet.* **2002**, *48*, 31−43.
(347) Gront, D.; Kmiecik, S.; Blaszczyk, M.; Ekonomiuk, D.; Kolinski, A. Optimization of protein models. *Wires Comput. Mol. Sci.* **2012**, *2*, 479−493.
(348) Porebski, P. J.; Cymborowski, M.; Pasenkiewicz-Gierula, M.; Minor, W. Fitmunk: improving protein structures by accurate, automatic modeling of side-chain conformations. *Acta Crystallogr. D Struct. Biol.* **2016**, *72*, 266−280.
(349) Gront, D.; Kolinski, A. BioShell–a package of tools for structural biology computations. *Bioinformatics* **2006**, *22*, 621−622.
(350) Gront, D.; Kolinski, A. Utility library for structural bioinformatics. *Bioinformatics* **2008**, *24*, 584−585.
(351) Eswar, N.; Webb, B.; Marti-Renom, M. A.; Madhusudhan, M. S.; Eramian, D.; Shen, M. Y.; Pieper, U.; Sali, A. Comparative protein structure modeling using MODELLER. *Curr. Protoc. Protein Sci.* **2007**, *Chapter 2*, Unit 2 9.10.1002/0471140864.ps0209s50
(352) Webb, B.; Sali, A. Protein structure modeling with MODELLER. *Methods Mol. Biol.* **2014**, *1137*, 1−15.
(353) Peterson, R. W.; Dutton, P. L.; Wand, A. J. Improved side-chain prediction accuracy using an ab initio potential energy function and a very large rotamer library. *Protein Sci.* **2004**, *13*, 735−751.
(354) Lu, M.; Dousis, A. D.; Ma, J. OPUS-Rota: a fast and accurate method for side-chain modeling. *Protein Sci.* **2008**, *17*, 1576−1585.
(355) Wallner, B.; Elofsson, A. All are not equal: a benchmark of different homology modeling programs. *Protein Sci.* **2005**, *14*, 1315−1327.
(356) Liang, S.; Zheng, D.; Zhang, C.; Standley, D. M. Fast and accurate prediction of protein side-chain conformations. *Bioinformatics* **2011**, *27*, 2913−2914.
(357) Moore, B. L.; Kelley, L. A.; Barber, J.; Murray, J. W.; MacDonald, J. T. High-quality protein backbone reconstruction from alpha carbons using Gaussian mixture models. *J. Comput. Chem.* **2013**, *34*, 1881−1889.
(358) Xiang, Z. X.; Honig, B. Extending the accuracy limits of prediction for side-chain conformations (vol 311, pg 421, 2001). *J. Mol. Biol.* **2001**, *312*, 419−419.
(359) Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(360) Li, Y.; Zhang, Y. REMO: A new protocol to refine full atomic protein models from C-alpha traces by optimizing hydrogen-bonding networks. *Proteins: Struct., Funct., Genet.* **2009**, *76*, 665−676.
(361) Maupetit, J.; Gautier, R.; Tuffery, P. SABBAC: online structural alphabet-based protein backbone reconstruction from alpha-carbon trace. *Nucleic Acids Res.* **2006**, *34*, W147−W151.
(362) Xu, J. B. Rapid protein side-chain packing via tree decomposition. *Research in Computational Molecular Biology, Proceedings* **2005**, *3500*, 423−439.
(363) Krivov, G. G.; Shapovalov, M. V.; Dunbrack, R. L., Jr. Improved prediction of protein side-chain conformations with SCWRL4. *Proteins: Struct., Funct., Genet.* **2009**, *77*, 778−795.
(364) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. Amber, a Package of Computer-Programs for Applying Molecular Mechanics, Normal-Mode Analysis, Molecular-Dynamics and Free-Energy Calculations to Simulate the Structural and Energetic Properties of Molecules. *Comput. Phys. Commun.* **1995**, *91*, 1−41.
(365) Skolnick, J.; Kolinski, A.; Ortiz, A. R. MONSSTER: a method for folding globular proteins with a small number of distance restraints. *J. Mol. Biol.* **1997**, *265*, 217−241.
(366) Wan, C. K.; Han, W.; Wu, Y. D. 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.
(367) 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.
(368) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E., 3rd; Laughton, C. A.; Orozco, M. Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys. J.* **2007**, *92*, 3817−3829.
(369) Dans, P. D.; Zeida, A.; Machado, M. R.; Pantano, S. A Coarse Grained Model for Atomic-Detailed DNA Simulations with Explicit Electrostatics. *J. Chem. Theory Comput.* **2010**, *6*, 1711−1725.
(370) Predeus, A. V.; Gul, S.; Gopal, S. M.; 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.
(371) Zacharias, M. Protein-protein docking with a reduced protein model accounting for side-chain flexibility. *Protein Sci.* **2003**, *12*, 1271−1282.
(372) Lin, Z. X.; Gunsteren, W. F. Refinement of the Application of the GROMOS 54A7 Force Field to beta-Peptides. *J. Comput. Chem.* **2013**, *34*, 2796−2805.
(373) Fiorucci, S.; Zacharias, M. Binding site prediction and improved scoring during flexible protein-protein docking with ATTRACT. *Proteins: Struct., Funct., Genet.* **2010**, *78*, 3131−3139.
(374) Maupetit, J.; Tuffery, P.; Derreumaux, P. A coarse-grained protein force field for folding and structure prediction. *Proteins: Struct., Funct., Genet.* **2007**, *69*, 394−408.
(375) Goga, N.; Melo, M. N.; Rzepiela, A. J.; de Vries, A. H.; Hadar, A.; Marrink, S. J.; Berendsen, H. J. Benchmark of Schemes for

DOI: 10.1021/acs.chemrev.6b00163
Chem. Rev. 2016, 116, 7898−7936

Multiscale Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2015**, *11*, 1389−1398.
(376) Nielsen, S. O.; Bulo, R. E.; Moore, P. B.; Ensing, B. Recent progress in adaptive multiscale molecular dynamics simulations of soft matter. *Phys. Chem. Chem. Phys.* **2010**, *12*, 12401−12414.
(377) Zhang, Y. Template-based modeling and free modeling by I-TASSER in CASP7. *Proteins: Struct., Funct., Genet.* **2007**, *69*, 108−117.
(378) Kim, D. E.; Chivian, D.; Baker, D. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res.* **2004**, *32*, W526−531.
(379) Kelley, L. A.; Mezulis, S.; Yates, C. M.; Wass, M. N.; Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **2015**, *10*, 845−858.
(380) Jefferys, B. R.; Kelley, L. A.; Sternberg, M. J. Protein folding requires crowd control in a simulated cell. *J. Mol. Biol.* **2010**, *397*, 1329−1338.
(381) Schindler, C. E.; de Vries, S. J.; Zacharias, M. Fully Blind Peptide-Protein Docking with pepATTRACT. *Structure* **2015**, *23*, 1507−1515.
(382) Schindler, C. E.; de Vries, S. J.; Zacharias, M. iATTRACT: simultaneous global and local interface optimization for protein-protein docking refinement. *Proteins: Struct., Funct., Genet.* **2015**, *83*, 248−258.
(383) Raveh, B.; London, N.; Zimmerman, L.; Schueler-Furman, O. Rosetta FlexPepDock ab-initio: simultaneous folding, docking and refinement of peptides onto their receptors. *PLoS One* **2011**, *6*, e18934.
(384) Jamroz, M.; Kolinski, A.; Kmiecik, S. CABS-flex predictions of protein flexibility compared with NMR ensembles. *Bioinformatics* **2014**, *30*, 2150−2154.
(385) Kruger, D. M.; Ahmed, A.; Gohlke, H. NMSim web server: integrated approach for normal mode-based geometric simulations of biologically relevant conformational transitions in proteins. *Nucleic Acids Res.* **2012**, *40*, W310−316.
(386) Camps, J.; Carrillo, O.; Emperador, A.; Orellana, L.; Hospital, A.; Rueda, M.; Cicin-Sain, D.; D'Abramo, M.; Gelpi, J. L.; Orozco, M. FlexServ: an integrated tool for the analysis of protein flexibility. *Bioinformatics* **2009**, *25*, 1709−1710.
(387) de Vries, S. J.; Schindler, C. E.; Chauvot de Beauchene, I.; Zacharias, M. A web interface for easy flexible protein-protein docking with ATTRACT. *Biophys. J.* **2015**, *108*, 462−465.
(388) May, A.; Zacharias, M. Protein-protein docking in CAPRI using ATTRACT to account for global and local flexibility. *Proteins: Struct., Funct., Genet.* **2007**, *69*, 774−780.
(389) Zacharias, M. ATTRACT: protein-protein docking in CAPRI using a reduced protein model. *Proteins: Struct., Funct., Genet.* **2005**, *60*, 252−256.
(390) de Vries, S.; Zacharias, M. Flexible docking and refinement with a coarse-grained protein model using ATTRACT. *Proteins: Struct., Funct., Genet.* **2013**, *81*, 2167−2174.
(391) Dehouck, Y.; Kwasigroch, J. M.; Rooman, M.; Gilis, D. BeAtMuSiC: Prediction of changes in protein-protein binding affinity on mutations. *Nucleic Acids Res.* **2013**, *41*, W333−339.
(392) Dehouck, Y.; Gilis, D.; Rooman, M. A new generation of statistical potentials for proteins. *Biophys. J.* **2006**, *90*, 4010−4017.
(393) Dehouck, Y.; Grosfils, A.; Folch, B.; Gilis, D.; Bogaerts, P.; Rooman, M. Fast and accurate predictions of protein stability changes upon mutations using statistical potentials and neural networks: PoPMuSiC-2.0. *Bioinformatics* **2009**, *25*, 2537−2543.
(394) Viswanath, S.; Ravikant, D. V.; Elber, R. DOCK/PIERR: web server for structure prediction of protein-protein complexes. *Methods Mol. Biol.* **2014**, *1137*, 199−207.
(395) Ravikant, D. V.; Elber, R. PIE-efficient filters and coarse grained potentials for unbound protein-protein docking. *Proteins: Struct., Funct., Genet.* **2010**, *78*, 400−419.
(396) Viswanath, S.; Ravikant, D. V.; Elber, R. Improving ranking of models for protein complexes with side chain modeling and atomic potentials. *Proteins: Struct., Funct., Genet.* **2013**, *81*, 592−606.
(397) Frappier, V.; Chartier, M.; Najmanovich, R. J. ENCoM server: exploring protein conformational space and the effect of mutations on protein function and stability. *Nucleic Acids Res.* **2015**, *43*, W395−400.
(398) Lyskov, S.; Gray, J. J. The RosettaDock server for local protein-protein docking. *Nucleic Acids Res.* **2008**, *36*, W233−238.
(399) London, N.; Schueler-Furman, O. FunHunt: model selection based on energy landscape characteristics. *Biochem. Soc. Trans.* **2008**, *36*, 1418−1421.
(400) Sircar, A.; Kim, E. T.; Gray, J. J. RosettaAntibody: antibody variable region homology modeling server. *Nucleic Acids Res.* **2009**, *37*, W474−479.
(401) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; et al. Atomic-level characterization of the structural dynamics of proteins. *Science* **2010**, *330*, 341−346.
(402) Freddolino, P. L.; Liu, F.; Gruebele, M.; Schulten, K. Ten-microsecond molecular dynamics simulation of a fast-folding WW domain. *Biophys. J.* **2008**, *94*, L75−77.
(403) Fersht, A. R. From the first protein structures to our current knowledge of protein folding: delights and scepticisms. *Nat. Rev. Mol. Cell Biol.* **2008**, *9*, 650−654.
(404) Schaeffer, R. D.; Fersht, A.; Daggett, V. Combining experiment and simulation in protein folding: closing the gap for small model systems. *Curr. Opin. Struct. Biol.* **2008**, *18*, 4−9.
(405) Sułkowska, J. I.; Cieplak, M. Selection of Optimal Variants of Gō-Like Models of Proteins through Studies of Stretching. *Biophys. J.* **2008**, *95*, 3174−3191.
(406) Prieto, L.; Rey, A. Topology-based potentials and the study of the competition between protein folding and aggregation. *J. Chem. Phys.* **2009**, *130*, 115101.
(407) Takada, S.; Kanada, R.; Tan, C.; Terakawa, T.; Li, W.; Kenzaki, H. Modeling Structural Dynamics of Biomolecular Complexes by Coarse-Grained Molecular Simulations. *Acc. Chem. Res.* **2015**, *48*, 3026−3035.
(408) Zhang, Z.; Chan, H. S. Competition between native topology and nonnative interactions in simple and complex folding kinetics of natural and designed proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 2920−2925.
(409) Shan, B.; Eliezer, D.; Raleigh, D. P. The unfolded state of the C-terminal domain of the ribosomal protein L9 contains both native and non-native structure. *Biochemistry* **2009**, *48*, 4707−4719.
(410) Rothwarf, D. M.; Scheraga, H. A. Role of non-native aromatic and hydrophobic interactions in the folding of hen egg white lysozyme. *Biochemistry* **1996**, *35*, 13797−13807.
(411) Jana, B.; Morcos, F.; Onuchic, J. N. From structure to function: the convergence of structure based models and co-evolutionary information. *Phys. Chem. Chem. Phys.* **2014**, *16*, 6496−6507.
(412) Okazaki, K.; Koga, N.; Takada, S.; Onuchic, J. N.; Wolynes, P. G. Multiple-basin energy landscapes for large-amplitude conformational motions of proteins: Structure-based molecular dynamics simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 11844−11849.
(413) Li, W.; Wang, W.; Takada, S. Energy landscape views for interplays among folding, binding, and allostery of calmodulin domains. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 10550−10555.
(414) Koga, N.; Takada, S. Folding-based molecular simulations reveal mechanisms of the rotary motor F1-ATPase. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 5367−5372.
(415) Han, W.; Schulten, K. Characterization of folding mechanisms of Trp-cage and WW-domain by network analysis of simulations with a hybrid-resolution model. *J. Phys. Chem. B* **2013**, *117*, 13367−13377.
(416) Han, W.; Wan, C. K.; Jiang, F.; Wu, Y. D. PACE Force Field for Protein Simulations. 1. Full Parameterization of Version 1 and Verification. *J. Chem. Theory Comput.* **2010**, *6*, 3373−3389.
(417) Thorpe, I. F.; Zhou, J.; Voth, G. A. Peptide folding using multiscale coarse-grained models. *J. Phys. Chem. B* **2008**, *112*, 13079−13090.
(418) Hills, R. D., Jr.; Lu, L.; Voth, G. A. Multiscale coarse-graining of the protein energy landscape. *PLoS Comput. Biol.* **2010**, *6*, e1000827.

DOI: 10.1021/acs.chemrev.6b00163
Chem. Rev. 2016, 116, 7898−7936

(419) Adhikari, A. N.; Freed, K. F.; Sosnick, T. R. De novo prediction of protein folding pathways and structure using the principle of sequential stabilization. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 17442−17447.
(420) Adhikari, A. N.; Freed, K. F.; Sosnick, T. R. Simplified protein models: predicting folding pathways and structure using amino acid sequences. *Phys. Rev. Lett.* **2013**, *111*, 028103.
(421) Rodrigues, J. P.; Levitt, M.; Chopra, G. KoBaMIN: a knowledge-based minimization web server for protein structure refinement. *Nucleic Acids Res.* **2012**, *40*, W323−328.
(422) Zhou, H. X.; Rivas, G.; Minton, A. P. Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences. *Annu. Rev. Biophys.* **2008**, *37*, 375−397.
(423) Feig, M.; Sugita, Y. Reaching new levels of realism in modeling biological macromolecules in cellular environments. *J. Mol. Graphics Modell.* **2013**, *45*, 144−156.
(424) Derreumaux, P. Coarse-grained models for protein folding and aggregation. *Methods Mol. Biol.* **2013**, *924*, 585−600.
(425) Borgia, M. B.; Borgia, A.; Best, R. B.; Steward, A.; Nettels, D.; Wunderlich, B.; Schuler, B.; Clarke, J. Single-molecule fluorescence reveals sequence-specific misfolding in multidomain proteins. *Nature* **2011**, *474*, 662−665.
(426) Morriss-Andrews, A.; Shea, J. E. Computational studies of protein aggregation: methods and applications. *Annu. Rev. Phys. Chem.* **2015**, *66*, 643−666.
(427) Shental-Bechor, D.; Smith, M. T.; Mackenzie, D.; Broom, A.; Marcovitz, A.; Ghashut, F.; Go, C.; Bralha, F.; Meiering, E. M.; Levy, Y. Nonnative interactions regulate folding and switching of myristoylated protein. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 17839−17844.
(428) Lu, D.; Yang, C.; Liu, Z. How hydrophobicity and the glycosylation site of glycans affect protein folding and stability: a molecular dynamics simulation. *J. Phys. Chem. B* **2012**, *116*, 390−400.
(429) Chan, H. S.; Zhang, Z.; Wallin, S.; Liu, Z. Cooperativity, local-nonlocal coupling, and nonnative interactions: principles of protein folding from coarse-grained models. *Annu. Rev. Phys. Chem.* **2011**, *62*, 301−326.
(430) Qin, M.; Wang, W.; Thirumalai, D. Protein folding guides disulfide bond formation. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 11241−11246.
(431) Jewett, A. I.; Shea, J. E. Reconciling theories of chaperonin accelerated folding with experimental evidence. *Cell. Mol. Life Sci.* **2010**, *67*, 255−276.
(432) Lucent, D.; England, J.; Pande, V. Inside the chaperonin toolbox: theoretical and computational models for chaperonin mechanism. *Phys. Biol.* **2009**, *6*, 015003.
(433) Stan, G.; Lorimer, G. H.; Thirumalai, D.; Brooks, B. R. Coupling between allosteric transitions in GroEL and assisted folding of a substrate protein. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 8803−8808.
(434) Orozco, M.; Orellana, L.; Hospital, A.; Naganathan, A. N.; Emperador, A.; Carrillo, O.; Gelpi, J. L. Coarse-grained representation of protein flexibility. Foundations, successes, and shortcomings. *Adv. Protein Chem. Struct. Biol.* **2011**, *85*, 183−215.
(435) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. How Robust Are Protein Folding Simulations with Respect to Force Field Parameterization? *Biophys. J.* **2011**, *100*, L47−L49.
(436) Martin-Garcia, F.; Papaleo, E.; Gomez-Puertas, P.; Boomsma, W.; Lindorff-Larsen, K. Comparing Molecular Dynamics Force Fields in the Essential Subspace. *PLoS One* **2015**, *10*, e012111410.1371/journal.pone.0121114
(437) Palazzesi, F.; Prakash, M. K.; Bonomi, M.; Barducci, A. Accuracy of current all-atom force-fields in modeling protein disordered states. *J. Chem. Theory Comput.* **2015**, *11*, 2−7.
(438) Rueda, M.; Ferrer-Costa, C.; Meyer, T.; Perez, A.; Camps, J.; Hospital, A.; Gelpi, J. L.; Orozco, M. A consensus view of protein dynamics. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 796−801.
(439) Emperador, A.; Carrillo, O.; Rueda, M.; Orozco, M. Exploring the suitability of coarse-grained techniques for the representation of protein dynamics. *Biophys. J.* **2008**, *95*, 2127−2138.
(440) Emperador, A.; Meyer, T.; Orozco, M. United-Atom Discrete Molecular Dynamics of Proteins Using Physics-Based Potentials. *J. Chem. Theory Comput.* **2008**, *4*, 2001−2010.
(441) Bowman, G. R. Accurately modeling nanosecond protein dynamics requires at least microseconds of simulation. *J. Comput. Chem.* **2016**, *37*, 558−566.
(442) Bahar, I.; Rader, A. J. Coarse-grained normal mode analysis in structural biology. *Curr. Opin. Struct. Biol.* **2005**, *15*, 586−592.
(443) Ma, J. Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. *Structure* **2005**, *13*, 373−380.
(444) Lexa, K. W.; Carlson, H. A. Full protein flexibility is essential for proper hot-spot mapping. *J. Am. Chem. Soc.* **2011**, *133*, 200−202.
(445) Antunes, D. A.; Devaurs, D.; Kavraki, L. E. Understanding the challenges of protein flexibility in drug design. *Expert Opin. Drug Discovery* **2015**, *10*, 1301−1313.
(446) Alvarez-Garcia, D.; Barril, X. Relationship between Protein Flexibility and Binding: Lessons for Structure-Based Drug Design. *J. Chem. Theory Comput.* **2014**, *10*, 2608−2614.
(447) Zhang, J.; Minary, P.; Levitt, M. Multiscale natural moves refine macromolecules using single-particle electron microscopy projection images. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 9845−9850.
(448) Consortium, T. U. UniProt: a hub for protein information. *Nucleic Acids Res.* **2014**.
(449) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(450) Rost, B. Twilight zone of protein sequence alignments. *Protein Eng., Des. Sel.* **1999**, *12*, 85−94.
(451) Liang, S.; Zhang, C.; Zhou, Y. LEAP: highly accurate prediction of protein loop conformations by integrating coarse-grained sampling and optimized energy scores with all-atom refinement of backbone and side chains. *J. Comput. Chem.* **2014**, *35*, 335−341.
(452) Kmiecik, S.; Jamroz, M.; Kolinski, M. Structure prediction of the second extracellular loop in G-protein-coupled receptors. *Biophys. J.* **2014**, *106*, 2408−2416.
(453) Goldfeld, D. A.; Zhu, K.; Beuming, T.; Friesner, R. A. Loop prediction for a GPCR homology model: algorithms and results. *Proteins: Struct., Funct., Genet.* **2013**, *81*, 214−228.
(454) Kolinski, M.; Filipek, S. Study of a structurally similar kappa opioid receptor agonist and antagonist pair by molecular dynamics simulations. *J. Mol. Model.* **2010**, *16*, 1567−1576.
(455) Zhang, Y. Interplay of I-TASSER and QUARK for template-based and ab initio protein structure prediction in CASP10. *Proteins: Struct., Funct., Genet.* **2014**, *82* (Suppl 2), 175−187.
(456) Szilagyi, A.; Zhang, Y. Template-based structure modeling of protein-protein interactions. *Curr. Opin. Struct. Biol.* **2014**, *24*, 10−23.
(457) Iacoangeli, A.; Marcatili, P.; Tramontano, A. Exploiting homology information in nontemplate based prediction of protein structures. *J. Chem. Theory Comput.* **2015**, *11*, 5045−5051.
(458) Zhang, W.; Yang, J.; He, B.; Walker, S. E.; Zhang, H.; Govindarajoo, B.; Virtanen, J.; Xue, Z.; Shen, H. B.; Zhang, Y. Integration of QUARK and I-TASSER for Ab Initio Protein Structure Prediction in CASP11. *Proteins: Struct., Funct., Genet.* **2015**, n/a.
(459) Xu, D.; Zhang, Y. Ab Initio structure prediction for Escherichia coli: towards genome-wide protein structure modeling and fold assignment. *Sci. Rep.* **2013**, *3*, 1895.
(460) Thevenet, P.; Shen, Y.; Maupetit, J.; Guyon, F.; Derreumaux, P.; Tuffery, P. PEP-FOLD: an updated de novo structure prediction server for both linear and disulfide bonded cyclic peptides. *Nucleic Acids Res.* **2012**, *40*, W288−293.
(461) Thévenet, P.; Rey, J.; Moroy, G.; Tuffery, P. In *Computational Peptidology*; Zhou, P., Huang, J., Eds.; Springer: New York, 2015; Vol. *1268*.

DOI: 10.1021/acs.chemrev.6b00163
Chem. Rev. 2016, 116, 7898−7936

(462) Samudrala, R.; Xia, Y.; Huang, E.; Levitt, M. Ab initio protein structure prediction using a combined hierarchical approach. *Proteins: Struct., Funct., Genet.* **1999**, *37* (Suppl 3), 194−198.
(463) Kolinski, A.; Skolnick, J. Monte Carlo simulations of protein folding. II. Application to protein A, ROP, and crambin. *Proteins: Struct., Funct., Genet.* **1994**, *18*, 353−366.
(464) Kolinski, A.; Skolnick, J. Monte Carlo simulations of protein folding. I. Lattice model and interaction scheme. *Proteins: Struct., Funct., Genet.* **1994**, *18*, 338−352.
(465) Tai, C. H.; Bai, H.; Taylor, T. J.; Lee, B. Assessment of template-free modeling in CASP10 and ROLL. *Proteins: Struct., Funct., Genet.* **2014**, *82* (Suppl 2), 57−83.
(466) Kinch, L.; Yong Shi, S.; Cong, Q.; Cheng, H.; Liao, Y.; Grishin, N. V. CASP9 assessment of free modeling target predictions. *Proteins: Struct., Funct., Genet.* **2011**, *79* (Suppl 10), 59−73.
(467) Ben-David, M.; Noivirt-Brik, O.; Paz, A.; Prilusky, J.; Sussman, J. L.; Levy, Y. Assessment of CASP8 structure predictions for template free targets. *Proteins: Struct., Funct., Genet.* **2009**, *77* (59), 50−65.
(468) Xu, D.; Zhang, Y. Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. *Proteins: Struct., Funct., Genet.* **2012**, *80*, 1715−1735.
(469) Moult, J.; Fidelis, K.; Kryshtafovych, A.; Schwede, T.; Tramontano, A. Critical assessment of methods of protein structure prediction (CASP)−round x. *Proteins: Struct., Funct., Genet.* **2014**, *82* (Suppl 2), 1−6.
(470) Kinch, L. N.; Li, W.; Monastyrskyy, B.; Kryshtafovych, A.; Grishin, N. V. Evaluation of free modeling targets in CASP11 and ROLL. *Proteins: Struct., Funct., Genet.* **2015**, n/a.
(471) Kim, H.; Kihara, D. Protein structure prediction using residue- and fragment-environment potentials in CASP11. *Proteins: Struct., Funct., Genet.* **2015**, n/a.
(472) Feig, M.; Mirjalili, V. Protein structure refinement via molecular-dynamics simulations: What works and what does not? *Proteins: Struct., Funct., Genet.* **2015**, n/a.
(473) Berggard, T.; Linse, S.; James, P. Methods for the detection and analysis of protein-protein interactions. *Proteomics* **2007**, *7*, 2833−2842.
(474) Setny, P.; Bahadur, R. P.; Zacharias, M. Protein-DNA docking with a coarse-grained force field. *BMC Bioinf.* **2012**, *13*, 228.
(475) Uusitalo, J. J.; Ingolfsson, H. I.; Akhshi, P.; Tieleman, D. P.; Marrink, S. J. Martini Coarse-Grained Force Field: Extension to DNA. *J. Chem. Theory Comput.* **2015**, *11*, 3932−3945.
(476) Hori, N.; Takada, S. Coarse-Grained Structure-Based Model for RNA-Protein Complexes Developed by Fluctuation Matching. *J. Chem. Theory Comput.* **2012**, *8*, 3384−3394.
(477) Setny, P.; Zacharias, M. A coarse-grained force field for Protein-RNA docking. *Nucleic Acids Res.* **2011**, *39*, 9118−9129.
(478) Negami, T.; Shimizu, K.; Terada, T. Coarse-Grained Molecular Dynamics Simulations of Protein-Ligand Binding. *J. Comput. Chem.* **2014**, *35*, 1835−1845.
(479) Li, D.; Liu, M. S.; Ji, B.; Hwang, K.; Huang, Y. Coarse-grained molecular dynamics of ligands binding into protein: The case of HIV-1 protease inhibitors. *J. Chem. Phys.* **2009**, *130*, 215102.
(480) Prasanna, X.; Chattopadhyay, A.; Sengupta, D. Cholesterol modulates the dimer interface of the beta(2)-adrenergic receptor via cholesterol occupancy sites. *Biophys. J.* **2014**, *106*, 1290−1300.
(481) Zacharias, M. Combining coarse-grained nonbonded and atomistic bonded interactions for protein modeling. *Proteins: Struct., Funct., Genet.* **2013**, *81*, 81−92.
(482) Solernou, A.; Fernandez-Recio, J. pyDockCG: new coarse-grained potential for protein-protein docking. *J. Phys. Chem. B* **2011**, *115*, 6032−6039.
(483) de Vries, S. J.; Zacharias, M. ATTRACT-EM: a new method for the computational assembly of large molecular machines using cryo-EM maps. *PLoS One* **2012**, *7*, e49733.
(484) Periole, X.; Knepp, A. M.; Sakmar, T. P.; Marrink, S. J.; Huber, T. Structural determinants of the supramolecular organization of G protein-coupled receptors in bilayers. *J. Am. Chem. Soc.* **2012**, *134*, 10959−10965.
(485) Steczkiewicz, K.; Zimmermann, M. T.; Kurcinski, M.; Lewis, B. A.; Dobbs, D.; Kloczkowski, A.; Jernigan, R. L.; Kolinski, A.; Ginalski, K. Human telomerase model shows the role of the TEN domain in advancing the double helix for the next polymerization step. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 9443−9448.
(486) Liu, S.; Vakser, I. A. DECK: Distance and environment-dependent, coarse-grained, knowledge-based potentials for protein-protein docking. *BMC Bioinf.* **2011**, *12*, 280.
(487) Ghoorah, A. W.; Devignes, M. D.; Smail-Tabbone, M.; Ritchie, D. W. Spatial clustering of protein binding sites for template based protein docking. *Bioinformatics* **2011**, *27*, 2820−2827.
(488) Andreani, J.; Faure, G.; Guerois, R. InterEvScore: a novel coarse-grained interface scoring function using a multi-body statistical potential coupled to evolution. *Bioinformatics* **2013**, *29*, 1742−1749.
(489) Solernou, A.; Fernandez-Recio, J. pyDockCG: new coarse-grained potential for protein-protein docking. *J. Phys. Chem. B* **2011**, *115*, 6032−6039.
(490) Frembgen-Kesner, T.; Elcock, A. H. Absolute protein-protein association rate constants from flexible, coarse-grained Brownian dynamics simulations: the role of intermolecular hydrodynamic interactions in barnase-barstar association. *Biophys. J.* **2010**, *99*, L75−77.
(491) Emperador, A.; Sfriso, P.; Villarreal, M. A.; Gelpi, J. L.; Orozco, M. PACSAB: Coarse-Grained Force Field for the Study of Protein-Protein Interactions and Conformational Sampling in Multiprotein Systems. *J. Chem. Theory Comput.* **2015**, *11*, 5929−5938.
(492) May, A.; Pool, R.; van Dijk, E.; Bijlard, J.; Abeln, S.; Heringa, J.; Feenstra, K. A. Coarse-grained versus atomistic simulations: realistic interaction free energies for real proteins. *Bioinformatics* **2014**, *30*, 326−334.
(493) Bastard, K.; Prevost, C.; Zacharias, M. Accounting for loop flexibility during protein-protein docking. *Proteins: Struct., Funct., Genet.* **2006**, *62*, 956−969.
(494) Schneider, S.; Saladin, A.; Fiorucci, S.; Prevost, C.; Zacharias, M. ATTRACT and PTools: open source programs for protein-protein docking. *Methods Mol. Biol.* **2012**, *819*, 221−232.
(495) Sacquin-Mora, S.; Laforet, E.; Lavery, R. Locating the active sites of enzymes using mechanical properties. *Proteins: Struct., Funct., Genet.* **2007**, *67*, 350−359.
(496) Poulain, P.; Saladin, A.; Hartmann, B.; Prevost, C. Insights on protein-DNA recognition by coarse grain modelling. *J. Comput. Chem.* **2008**, *29*, 2582−2592.
(497) Saladin, A.; Amourda, C.; Poulain, P.; Ferey, N.; Baaden, M.; Zacharias, M.; Delalande, O.; Prevost, C. Modeling the early stage of DNA sequence recognition within RecA nucleoprotein filaments. *Nucleic Acids Res.* **2010**, *38*, 6313−6323.
(498) Molza, A. E.; Ferey, N.; Czjzek, M.; Le Rumeur, E.; Hubert, J. F.; Tek, A.; Laurent, B.; Baaden, M.; Delalande, O. Innovative interactive flexible docking method for multi-scale reconstruction elucidates dystrophin molecular assembly. *Faraday Discuss.* **2014**, *169*, 45−62.
(499) Lensink, M. F.; Wodak, S. J. Docking, scoring, and affinity prediction in CAPRI. *Proteins: Struct., Funct., Genet.* **2013**, *81*, 2082−2095.
(500) Kozakov, D.; Beglov, D.; Bohnuud, T.; Mottarella, S. E.; Xia, B.; Hall, D. R.; Vajda, S. How good is automated protein docking? *Proteins: Struct., Funct., Genet.* **2013**, *81*, 2159−2166.
(501) Vajda, S.; Kozakov, D. Convergence and combination of methods in protein-protein docking. *Curr. Opin. Struct. Biol.* **2009**, *19*, 164−170.
(502) Huang, S. Y. Search strategies and evaluation in protein-protein docking: principles, advances and challenges. *Drug Discovery Today* **2014**, *19*, 1081−1096.
(503) Huang, S. Y. Exploring the potential of global protein-protein docking: an overview and critical assessment of current programs for automatic ab initio docking. *Drug Discovery Today* **2015**, *20*, 969−977.
(504) Tuffery, P.; Derreumaux, P. Flexibility and binding affinity in protein-ligand, protein-protein and multi-component protein inter-

DOI: 10.1021/acs.chemrev.6b00163
Chem. Rev. 2016, 116, 7898−7936

actions: limitations of current computational approaches. *J. R. Soc., Interface* **2012**, *9*, 20−33.
(505) Li, H.; Lu, L.; Chen, R.; Quan, L.; Xia, X.; Lu, Q. PaFlexPepDock: parallel ab-initio docking of peptides onto their receptors with full flexibility based on Rosetta. *PLoS One* **2014**, *9*, e94769.
(506) Wabik, J.; Kurcinski, M.; Kolinski, A. Coarse-Grained Modeling of Peptide Docking Associated with Large Conformation Transitions of the Binding Protein: Troponin I Fragment-Troponin C System. *Molecules* **2015**, *20*, 10763−10780.
(507) Ravikumar, K. M.; Huang, W.; Yang, S. Coarse-grained simulations of protein-protein association: an energy landscape perspective. *Biophys. J.* **2012**, *103*, 837−845.
(508) Irback, A.; Jonsson, S. A. E.; Linnemann, N.; Linse, B.; Wallin, S. Aggregate geometry in amyloid fibril nucleation. *Phys. Rev. Lett.* **2013**, *110*, 058101.
(509) Bieler, N. S.; Knowles, T. P.; Frenkel, D.; Vacha, R. Connecting macroscopic observables and microscopic assembly events in amyloid formation using coarse grained simulations. *PLoS Comput. Biol.* **2012**, *8*, e1002692.
(510) Li, M. S.; Klimov, D. K.; Straub, J. E.; Thirumalai, D. Probing the mechanisms of fibril formation using lattice models. *J. Chem. Phys.* **2008**, *129*, 175101.
(511) Friedman, R.; Caflisch, A. Surfactant effects on amyloid aggregation kinetics. *J. Mol. Biol.* **2011**, *414*, 303−312.
(512) Bellesia, G.; Shea, J. E. Self-assembly of beta-sheet forming peptides into chiral fibrillar aggregates. *J. Chem. Phys.* **2007**, *126*, 245104.
(513) Carmichael, S. P.; Shell, M. S. A new multiscale algorithm and its application to coarse-grained peptide models for self-assembly. *J. Phys. Chem. B* **2012**, *116*, 8383−8393.
(514) Cheon, M.; Chang, I.; Hall, C. K. Extending the PRIME model for protein aggregation to all 20 amino acids. *Proteins: Struct., Funct., Genet.* **2010**, *78*, 2950−2960.
(515) Uversky, V. N. Introduction to intrinsically disordered proteins (IDPs). *Chem. Rev.* **2014**, *114*, 6557−6560.
(516) Habchi, J.; Tompa, P.; Longhi, S.; Uversky, V. N. Introducing protein intrinsic disorder. *Chem. Rev.* **2014**, *114*, 6561−6588.
(517) Mittag, T.; Kay, L. E.; Forman-Kay, J. D. Protein dynamics and conformational disorder in molecular recognition. *J. Mol. Recognit.* **2009**, *23*, 105−116.
(518) Uversky, V. N.; Gillespie, J. R.; Fink, A. L. Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins: Struct., Funct., Genet.* **2000**, *41*, 415−427.
(519) Wright, P. E.; Dyson, H. J. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **1999**, *293*, 321−331.
(520) Zhang, Y.; Stec, B.; Godzik, A. Between order and disorder in protein structures: analysis of "dual personality" fragments in proteins. *Structure* **2007**, *15*, 1141−1147.
(521) Eliezer, D. Biophysical characterization of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **2009**, *19*, 23−30.
(522) Rauscher, S.; Pomès, R. Molecular simulations of protein disorderThis paper is one of a selection of papers published in this special issue entitled "Canadian Society of Biochemistry, Molecular & Cellular Biology 52nd Annual Meeting — Protein Folding: Principles and Diseases" and has undergone the Journal's usual peer review process. *Biochem. Cell Biol.* **2010**, *88*, 269−290.
(523) Baker, C. M.; Best, R. B. Insights into the binding of intrinsically disordered proteins from molecular dynamics simulation. *Wires Comput. Mol. Sci.* **2014**, *4*, 182−198.
(524) David de, S.; Christopher, M. B.; Robert, B. B. In *Computational Approaches to Protein Dynamics*; CRC Press: Boca Raton, FL, 2014.
(525) Rauscher, S.; Gapsys, V.; Gajda, M. J.; Zweckstetter, M.; de Groot, B. L.; Grubmuller, H. Structural Ensembles of Intrinsically Disordered Proteins Depend Strongly on Force Field: A Comparison to Experiment. *J. Chem. Theory Comput.* **2015**, *11*, 5513−5524.
(526) Sugase, K.; Dyson, H. J.; Wright, P. E. Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* **2007**, *447*, 1021−1025.
(527) Chen, H. F. Molecular dynamics simulation of phosphorylated KID post-translational modification. *PLoS One* **2009**, *4*, e651610.1371/journal.pone.0006516
(528) Huang, Y.; Liu, Z. Kinetic Advantage of Intrinsically Disordered Proteins in Coupled Folding−Binding Process: A Critical Assessment of the "Fly-Casting" Mechanism. *J. Mol. Biol.* **2009**, *393*, 1143−1159.
(529) Ganguly, D.; Chen, J. Topology-based modeling of intrinsically disordered proteins: Balancing intrinsic folding and intermolecular interactions. *Proteins: Struct., Funct., Genet.* **2011**, *79*, 1251−1266.
(530) Huang, Y.; Liu, Z. Nonnative interactions in coupled folding and binding processes of intrinsically disordered proteins. *PLoS One* **2010**, *5*, e15375.
(531) Turjanski, A. G.; Gutkind, J. S.; Best, R. B.; Hummer, G. Binding-induced folding of a natively unstructured transcription factor. *PLoS Comput. Biol.* **2008**, *4*, e1000060.
(532) Huang, Y.; Liu, Z. Smoothing molecular interactions: the "kinetic buffer" effect of intrinsically disordered proteins. *Proteins: Struct., Funct., Genet.* **2010**, *78*, 3251−3259.
(533) Law, S. M.; Ahlstrom, L. S.; Panahi, A.; Brooks, C. L. Hamiltonian Mapping Revisited: Calibrating Minimalist Models to Capture Molecular Recognition by Intrinsically Disordered Proteins. *J. Phys. Chem. Lett.* **2014**, *5*, 3441−3444.
(534) Knott, M.; Best, R. B. Discriminating binding mechanisms of an intrinsically disordered protein via a multi-state coarse-grained model. *J. Chem. Phys.* **2014**, *140*, 175102.
(535) Rutter, G. O.; Brown, A. H.; Quigley, D.; Walsh, T. R.; Allen, M. P. Testing the transferability of a coarse-grained model to intrinsically disordered proteins. *Phys. Chem. Chem. Phys.* **2015**, *17*, 31741−31749.
(536) Terakawa, T.; Higo, J.; Takada, S. Multi-scale ensemble modeling of modular proteins with intrinsically disordered linker regions: application to p53. *Biophys. J.* **2014**, *107*, 721−729.
(537) Norgaard, A. B.; Ferkinghoff-Borg, J.; Lindorff-Larsen, K. Experimental Parameterization of an Energy Function for the Simulation of Unfolded Proteins. *Biophys. J.* **2008**, *94*, 182−192.
(538) Amorós, D.; Ortega, A.; García de la Torre, J. Prediction of Hydrodynamic and Other Solution Properties of Partially Disordered Proteins with a Simple, Coarse-Grained Model. *J. Chem. Theory Comput.* **2013**, *9*, 1678−1685.
(539) Qin, S.; Zhou, H.-X. Effects of Macromolecular Crowding on the Conformational Ensembles of Disordered Proteins. *J. Phys. Chem. Lett.* **2013**, *4*, 3429−3434.
(540) Wang, J.; Wang, Y.; Chu, X.; Hagen, S. J.; Han, W.; Wang, E. Multi-Scaled Explorations of Binding-Induced Folding of Intrinsically Disordered Protein Inhibitor IA3 to its Target Enzyme. *PLoS Comput. Biol.* **2011**, *7*, e1001118.
(541) Stadler, A. M.; Stingaciu, L.; Radulescu, A.; Holderer, O.; Monkenbusch, M.; Biehl, R.; Richter, D. Internal Nanosecond Dynamics in the Intrinsically Disordered Myelin Basic Protein. *J. Am. Chem. Soc.* **2014**, *136*, 6987−6994.
(542) Yu, H.; Han, W.; Ma, W.; Schulten, K. Transient beta-hairpin formation in alpha-synuclein monomer revealed by coarse-grained molecular dynamics simulation. *J. Chem. Phys.* **2015**, *143*, 243142.
(543) Ghavami, A.; Veenhoff, L. M.; van der Giessen, E.; Onck, P. R. Probing the disordered domain of the nuclear pore complex through coarse-grained molecular dynamics simulations. *Biophys. J.* **2014**, *107*, 1393−1402.
(544) Ahlstrom, L. S.; Law, S. M.; Dickson, A.; Brooks, C. L. Multiscale Modeling of a Conditionally Disordered pH-Sensing Chaperone. *J. Mol. Biol.* **2015**, *427*, 1670−1680.
(545) Koehler Leman, J.; Ulmschneider, M. B.; Gray, J. J. Computational modeling of membrane proteins. *Proteins: Struct., Funct., Genet.* **2015**, *83*, 1−24.
(546) Baker, M. Making membrane proteins for structures: a trillion tiny tweaks. *Nat. Methods* **2010**, *7*, 429−434.

DOI: 10.1021/acs.chemrev.6b00163
Chem. Rev. 2016, 116, 7898−7936

(547) Marrink, S. J.; Mark, A. E. Molecular dynamics simulation of the formation, structure, and dynamics of small phospholipid vesicles. *J. Am. Chem. Soc.* **2003**, *125*, 15233−15242.
(548) Knecht, V.; Marrink, S. J. Molecular dynamics simulations of lipid vesicle fusion in atomic detail. *Biophys. J.* **2007**, *92*, 4254−4261.
(549) Bennett, W. F.; MacCallum, J. L.; Hinner, M. J.; Marrink, S. J.; Tieleman, D. P. Molecular view of cholesterol flip-flop and chemical potential in different membrane environments. *J. Am. Chem. Soc.* **2009**, *131*, 12714−12720.
(550) Ogushi, F.; Ishitsuka, R.; Kobayashi, T.; Sugita, Y. Rapid flip-flop motions of diacylglycerol and ceramide in phospholipid bilayers. *Chem. Phys. Lett.* **2012**, *522*, 96.
(551) Esteban-Martin, S.; Risselada, H. J.; Salgado, J.; Marrink, S. J. Stability of asymmetric lipid bilayers assessed by molecular dynamics simulations. *J. Am. Chem. Soc.* **2009**, *131*, 15194−15202.
(552) Kasson, P. M.; Lindahl, E.; Pande, V. S. Atomic-resolution simulations predict a transition state for vesicle fusion defined by contact of a few lipid tails. *PLoS Comput. Biol.* **2010**, *6*, e1000829.
(553) Laganowsky, A.; Reading, E.; Allison, T. M.; Ulmschneider, M. B.; Degiacomi, M. T.; Baldwin, A. J.; Robinson, C. V. Membrane proteins bind lipids selectively to modulate their structure and function. *Nature* **2014**, *510*, 172.
(554) Bond, P. J.; Sansom, M. S. Insertion and assembly of membrane proteins via simulation. *J. Am. Chem. Soc.* **2006**, *128*, 2697−2704.
(555) Hall, B. A.; Chetwynd, A. P.; Sansom, M. S. Exploring peptide-membrane interactions with coarse-grained MD simulations. *Biophys. J.* **2011**, *100*, 1940−1948.
(556) Sobhanifar, S.; Schneider, B.; Lohr, F.; Gottstein, D.; Ikeya, T.; Mlynarczyk, K.; Pulawski, W.; Ghoshdastider, U.; Kolinski, M.; Filipek, S.; et al. Structural investigation of the C-terminal catalytic fragment of presenilin 1. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 9644−9649.
(557) Klingelhoefer, J. W.; Carpenter, T.; Sansom, M. S. Peptide nanopores and lipid bilayers: interactions by coarse-grained molecular-dynamics simulations. *Biophys. J.* **2009**, *96*, 3519−3528.
(558) Balali-Mood, K.; Bond, P. J.; Sansom, M. S. Interaction of monotopic membrane enzymes with a lipid bilayer: a coarse-grained MD simulation study. *Biochemistry* **2009**, *48*, 2135−2145.
(559) Sansom, M. S.; Scott, K. A.; Bond, P. J. Coarse-grained simulation: a high-throughput computational approach to membrane proteins. *Biochem. Soc. Trans.* **2008**, *36*, 27−32.
(560) Dominguez, L.; Meredith, S. C.; Straub, J. E.; Thirumalai, D. Transmembrane fragment structures of amyloid precursor protein depend on membrane surface curvature. *J. Am. Chem. Soc.* **2014**, *136*, 854−857.
(561) Long, S. B.; Tao, X.; Campbell, E. B.; MacKinnon, R. Atomic structure of a voltage-dependent K+ channel in a lipid membrane-like environment. *Nature* **2007**, *450*, 376−U373.
(562) Sengupta, D.; Chattopadhyay, A. Identification of cholesterol binding sites in the serotonin1A receptor. *J. Phys. Chem. B* **2012**, *116*, 12991−12996.
(563) Koivuniemi, A.; Vuorela, T.; Kovanen, P. T.; Vattulainen, I.; Hyvonen, M. T. Lipid exchange mechanism of the cholesteryl ester transfer protein clarified by atomistic and coarse-grained simulations. *PLoS Comput. Biol.* **2012**, *8*, e1002299.
(564) Pleskot, R.; Pejchar, P.; Zarsky, V.; Staiger, C. J.; Potocky, M. Structural insights into the inhibition of actin-capping protein by interactions with phosphatidic acid and phosphatidylinositol (4,5)-bisphosphate. *PLoS Comput. Biol.* **2012**, *8*, e1002765.
(565) Stansfeld, P. J.; Hopkinson, R.; Ashcroft, F. M.; Sansom, M. S. PIP(2)-binding site in Kir channels: definition by multiscale biomolecular simulations. *Biochemistry* **2009**, *48*, 10926−10933.
(566) Yin, F.; Kindt, J. T. Hydrophobic mismatch and lipid sorting near OmpA in mixed bilayers: atomistic and coarse-grained simulations. *Biophys. J.* **2012**, *102*, 2279−2287.
(567) Arnarez, C.; Mazat, J. P.; Elezgaray, J.; Marrink, S. J.; Periole, X. Evidence for cardiolipin binding sites on the membrane-exposed surface of the cytochrome bc1. *J. Am. Chem. Soc.* **2013**, *135*, 3112−3120.
(568) Weingarth, M.; Prokofyev, A.; van der Cruijsen, E. A.; Nand, D.; Bonvin, A. M.; Pongs, O.; Baldus, M. Structural determinants of specific lipid binding to potassium channels. *J. Am. Chem. Soc.* **2013**, *135*, 3983−3988.
(569) Sharma, S.; Juffer, A. H. An atomistic model for assembly of transmembrane domain of T cell receptor complex. *J. Am. Chem. Soc.* **2013**, *135*, 2188−2197.
(570) Dominguez, L.; Foster, L.; Meredith, S. C.; Straub, J. E.; Thirumalai, D. Structural heterogeneity in transmembrane amyloid precursor protein homodimer is a consequence of environmental selection. *J. Am. Chem. Soc.* **2014**, *136*, 9619−9626.
(571) Carpenter, T.; Bond, P. J.; Khalid, S.; Sansom, M. S. Self-assembly of a simple membrane protein: coarse-grained molecular dynamics simulations of the influenza M2 channel. *Biophys. J.* **2008**, *95*, 3790−3801.
(572) Provasi, D.; Boz, M. B.; Johnston, J. M.; Filizola, M. Preferred supramolecular organization and dimer interfaces of opioid receptors from simulated self-association. *PLoS Comput. Biol.* **2015**, *11*, e1004148.
(573) Johnston, J. M.; Wang, H.; Provasi, D.; Filizola, M. Assessing the relative stability of dimer interfaces in g protein-coupled receptors. *PLoS Comput. Biol.* **2012**, *8*, e1002649.
(574) Ayton, G. S.; Voth, G. A. Multiscale simulation of protein mediated membrane remodeling. *Semin. Cell Dev. Biol.* **2010**, *21*, 357−362.
(575) Genheden, S.; Essex, J. W. A simple and transferable all-atom/coarse-grained hybrid model to study membrane processes. *J. Chem. Theory Comput.* **2015**, *11*, 4749−4759.
(576) Vorobyov, I.; Kim, I.; Chu, Z. T.; Warshel, A. Refining the treatment of membrane proteins by coarse-grained models. *Proteins: Struct., Funct., Genet.* **2016**, *84*, 92−117.
(577) Kessel, A.; Shental-Bechor, D.; Haliloglu, T.; Ben-Tal, N. Interactions of hydrophobic peptides with lipid bilayers: Monte Carlo simulations with M2delta. *Biophys. J.* **2003**, *85*, 3431−3444.
(578) Maddox, M. W.; Longo, M. L. A Monte Carlo study of peptide insertion into lipid bilayers: equilibrium conformations and insertion mechanisms. *Biophys. J.* **2002**, *82*, 244−263.
(579) Milik, M.; Skolnick, J. Insertion of peptide chains into lipid membranes: an off-lattice Monte Carlo dynamics model. *Proteins: Struct., Funct., Genet.* **1993**, *15*, 10−25.
(580) Baumgartner, A. Insertion and hairpin formation of membrane proteins: a Monte Carlo study. *Biophys. J.* **1996**, *71*, 1248−1255.
(581) Lopez, C. F.; Nielsen, S. O.; Srinivas, G.; Degrado, W. F.; Klein, M. L. Probing Membrane Insertion Activity of Antimicrobial Polymers via Coarse-grain Molecular Dynamics. *J. Chem. Theory Comput.* **2006**, *2*, 649−655.
(582) Dryga, A.; Chakrabarty, S.; Vicatos, S.; Warshel, A. Realistic simulation of the activation of voltage-gated ion channels. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 3335−3340.
(583) Kim, I.; Warshel, A. Coarse-grained simulations of the gating current in the voltage-activated Kv1.2 channel. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 2128−2133.
(584) Rychkova, A.; Vicatos, S.; Warshel, A. On the energetics of translocon-assisted insertion of charged transmembrane helices into membranes. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 17598−17603.
(585) Alford, R. F.; Koehler Leman, J.; Weitzner, B. D.; Duran, A. M.; Tilley, D. C.; Elazar, A.; Gray, J. J. An Integrated Framework Advancing Membrane Protein Modeling and Design. *PLoS Comput. Biol.* **2015**, *11*, e1004398.
(586) Xu, X.; Yan, C.; Wohlhueter, R.; Ivanov, I. Integrative Modeling of Macromolecular Assemblies from Low to Near-Atomic Resolution. *Comput. Struct. Biotechnol. J.* **2015**, *13*, 492−503.
(587) Latek, D.; Kolinski, A. CABS-NMR–De novo tool for rapid global fold determination from chemical shifts, residual dipolar couplings and sparse methyl-methyl NOEs. *J. Comput. Chem.* **2011**, *32*, 536−544.
(588) van der Schot, G.; Bonvin, A. M. Performance of the WeNMR CS-Rosetta3 web server in CASD-NMR. *J. Biomol. NMR* **2015**, *62*, 497−502.

DOI: 10.1021/acs.chemrev.6b00163
Chem. Rev. 2016, 116, 7898−7936

(589) Rossi, P.; Shi, L.; Liu, G.; Barbieri, C. M.; Lee, H. W.; Grant, T. D.; Luft, J. R.; Xiao, R.; Acton, T. B.; Snell, E. H.; et al. A hybrid NMR/SAXS-based approach for discriminating oligomeric protein interfaces using Rosetta. *Proteins: Struct., Funct., Genet.* **2015**, *83*, 309−317.
(590) Jang, R.; Wang, Y.; Xue, Z. D.; Zhang, Y. NMR data-driven structure determination using NMR-I-TASSER in the CASD-NMR experiment. *J. Biomol. NMR* **2015**, *62*, 511−525.
(591) Zheng, W. Accurate flexible fitting of high-resolution protein structures into cryo-electron microscopy maps using coarse-grained pseudo-energy minimization. *Biophys. J.* **2011**, *100*, 478−488.
(592) Grubisic, I.; Shokhirev, M. N.; Orzechowski, M.; Miyashita, O.; Tama, F. Biased coarse-grained molecular dynamics simulation approach for flexible fitting of X-ray structure into cryo electron microscopy maps. *J. Struct. Biol.* **2010**, *169*, 95−105.
(593) Zhang, J.; Baker, M. L.; Schroder, G. F.; Douglas, N. R.; Reissmann, S.; Jakana, J.; Dougherty, M.; Fu, C. J.; Levitt, M.; Ludtke, S. J.; et al. Mechanism of folding chamber closure in a group II chaperonin. *Nature* **2010**, *463*, 379−383.
(594) DiMaio, F.; Song, Y.; Li, X.; Brunner, M. J.; Xu, C.; Conticello, V.; Egelman, E.; Marlovits, T. C.; Cheng, Y.; Baker, D. Atomic-accuracy models from 4.5-A cryo-electron microscopy data with density-guided iterative local refinement. *Nat. Methods* **2015**, *12*, 361−365.
(595) Demers, J. P.; Habenstein, B.; Loquet, A.; Kumar Vasa, S.; Giller, K.; Becker, S.; Baker, D.; Lange, A.; Sgourakis, N. G. High-resolution structure of the Shigella type-III secretion needle by solid-state NMR and cryo-electron microscopy. *Nat. Commun.* **2014**, *5*, 4976.
(596) Esquivel-Rodriguez, J.; Kihara, D. Computational methods for constructing protein structure models from 3D electron microscopy maps. *J. Struct. Biol.* **2013**, *184*, 93−102.
(597) Shrestha, R.; Berenger, F.; Zhang, K. Y. Accelerating ab initio phasing with de novo models. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2011**, *67*, 804−812.
(598) Shrestha, R.; Simoncini, D.; Zhang, K. Y. Error-estimation-guided rebuilding of de novo models increases the success rate of ab initio phasing. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2012**, *68*, 1522−1534.
(599) Yang, S.; Park, S.; Makowski, L.; Roux, B. A rapid coarse residue-based computational method for x-ray solution scattering characterization of protein folds and multiple conformational states of large protein complexes. *Biophys. J.* **2009**, *96*, 4449−4463.
(600) Stovgaard, K.; Andreetta, C.; Ferkinghoff-Borg, J.; Hamelryck, T. Calculation of accurate small angle X-ray scattering curves from coarse-grained protein models. *BMC Bioinf.* **2010**, *11*, 429.
(601) Hagita, K.; McGreevy, R. L.; Arai, T.; Inui, M.; Matsuda, K.; Tamura, K. First example of multi-scale reverse Monte Carlo modeling for small-angle scattering experimental data using reverse mapping from coarse-grained particles to atoms. *J. Phys.: Condens. Matter* **2010**, *22*, 404215.
(602) Haxton, T. K. High-Resolution Coarse-Grained Modeling Using Oriented Coarse-Grained Sites. *J. Chem. Theory Comput.* **2015**, *11*, 1244−1254.
(603) Kang, M.; Roberts, C.; Cheng, Y.; Chang, C. E. Gating and Intermolecular Interactions in Ligand-Protein Association: Coarse-Grained Modeling of HIV-1 Protease. *J. Chem. Theory Comput.* **2011**, *7*, 3438−3446.
(604) Negami, T.; Shimizu, K.; Terada, T. Protein-Ligand Binding Simulation with the Martini Coarse-Grained Force Field. *Biophys. J.* **2014**, *106*, 609a.10.1016/j.bpj.2013.11.3370
(605) Ruiz-Herrero, T.; Estrada, J.; Guantes, R.; Miguez, D. G. A tunable coarse-grained model for ligand-receptor interaction. *PLoS Comput. Biol.* **2013**, *9*, e1003274.
(606) de Jong, D. H.; Baoukina, S.; Ingolfsson, H. I.; Marrink, S. J. Martini straight: Boosting performance using a shorter cutoff and GPUs. *Comput. Phys. Commun.* **2016**, *199*, 1−7.
(607) Lettieri, S.; Zuckerman, D. M. Accelerating molecular Monte Carlo simulations using distance and orientation-dependent energy tables: tuning from atomistic accuracy to smoothed "coarse-grained" models. *J. Comput. Chem.* **2012**, *33*, 268−275.
(608) Tunbridge, I.; Best, R. B.; Gain, J.; Kuttel, M. M. Simulation of Coarse-Grained Protein-Protein Interactions with Graphics Processing Units. *J. Chem. Theory Comput.* **2010**, *6*, 3588−3600.
(609) Czaplewski, C.; Liwo, A.; Makowski, M.; Oldziej, S.; Scheraga, H. A. Coarse-Grained Models of Proteins: Theory and Applications. *Multiscale Approaches to Protein Modeling* **2011**, 35−83.
(610) Thorpe, I. F.; Goldenberg, D. P.; Voth, G. A. Exploration of transferability in multiscale coarse-grained peptide models. *J. Phys. Chem. B* **2011**, *115*, 11911−11926.
(611) Potoyan, D. A.; Savelyev, A.; Papoian, G. A. Recent successes in coarse-grained modeling of DNA. *Wires Comput. Mol. Sci.* **2013**, *3*, 69−83.
(612) Boniecki, M. J.; Lach, G.; Dawson, W. K.; Tomala, K.; Lukasz, P.; Soltysinski, T.; Rother, K. M.; Bujnicki, J. M. SimRNA: a coarse-grained method for RNA folding simulations and 3D structure prediction. *Nucleic Acids Res.* **2016**, *44*, e63.
(613) Hyeon, C.; Denesyuk, N. A.; Thirumalai, D. Development and Applications of Coarse-Grained Models for RNA. *Isr. J. Chem.* **2014**, *54*, 1358−1373.
(614) Lopez, C. A.; Rzepiela, A. J.; de Vries, A. H.; Dijkhuizen, L.; Hunenberger, P. H.; Marrink, S. J.; Martini Coarse-Grained. Force Field: Extension to Carbohydrates. *J. Chem. Theory Comput.* **2009**, *5*, 3195−3210.
(615) Lopez, C. A.; Bellesia, G.; Redondo, A.; Langan, P.; Chundawat, S. P. S.; Dale, B. E.; Marrink, S. J.; Gnanakaran, S. MARTINI Coarse-Grained Model for Crystalline Cellulose Microfibers. *J. Phys. Chem. B* **2015**, *119*, 465−473.

DOI: 10.1021/acs.chemrev.6b00163
Chem. Rev. 2016, 116, 7898−7936