Multiscale modeling of macromolecular biosystems

Samuel C. Flores*, Julie Bernauer*, Seokmin Shin, Ruhong Zhou and Xuhui Huang*

Submitted: 1st September 2011; Received (in revised form): 12th December 2011

Abstract
In this article, we review the recent progress in multiresolution modeling of structure and dynamics of protein, RNA and their complexes. Many approaches using both physics-based and knowledge-based potentials have been developed at multiple granularities to model both protein and RNA. Coarse graining can be achieved not only in the length, but also in the time domain using discrete time and discrete state kinetic network models. Models with different resolutions can be combined either in a sequential or parallel fashion. Similarly, the modeling of assemblies is also often achieved using multiple granularities. The progress shows that a multiresolution approach has considerable potential to continue extending the length and time scales of macromolecular modeling.

Keywords: multiscale modeling; protein structure and dynamics; nucleic acid modeling; protein assemblies

INTRODUCTION
Modeling the folding and assembly of biological macromolecules is challenging, since it is at large size and long time scales that require great computational cost. In response, many methods have been developed to reduce the computational cost by coarsening the granularity of the problem. Inevitably, the accuracy is lower for these methods and thus, in recent years, a consensus has emerged that one should work at multiple levels of resolution, simultaneously or in parallel [1] in a multiresolution approach.

At the coarsest granularity, an entire macromolecule can be treated as a single particle, whereas at the finest level each atom can be treated as a separate entity; indeed even higher levels of theory are possible [2]. There are many grainings in between, for example the kinematic or force unit may be a domain, a secondary structure element, a residue or chemical groups such as bases and ribose rings [3]. A simulation can be carried out at any level of resolution, or different levels can be used for the same problem either in serial or in parallel [1]. Furthermore, the force field and the kinematics can be treated at different levels of granularity [4], for example, secondary structure elements can be rigidified while interelement interactions can be treated at the all-atom level of resolution. The methods may consist of force fields [5, 6], sampling, structure prediction [5, 6] and dynamics tools [4], as well as novel algorithms.

Given the limited space, this mini-review is not aimed to be a complete review of this emerging field,
but instead to select a few specific examples in the following three main categories, some from our own studies, to illustrate the concept of multiscale modeling and its applications in macromolecular biological systems:

- protein structure and dynamics;
- strategies for the modeling of nucleic acids;
- interactions, complexes and assemblies.

## MULTISCALE MODELING OF PROTEIN STRUCTURE AND DYNAMICS

Protein modeling is challenging because of the wide range of length and time spans. Modeling biological processes such as membrane pore assembly requires accurate description at both atomic/molecular scale (membrane protein) and mesoscopic/macrosopic scale (lipids) [7]. Furthermore, many biological processes of interest occur on a timescale (typically microseconds to milliseconds) that is much longer than what atomistic molecular dynamics (MD) simulations can reach routinely (typically hundreds of nanoseconds to microseconds). Therefore, it is computationally intractable to model complex protein systems in atomic detail, while also reaching biologically relevant timescales.

Multiscale approaches address the above challenge by coarse graining in either the length or the time domain. Simplifying protein representations in the length scale is a natural and popular solution to bridge the timescale gap [1, 8–10]. In such simulations, different length-scale resolutions can be mixed either in a sequential or parallel fashion [1]. The other solution for bridging the timescale gap is to construct models from short atomistic simulations to predict long timescale dynamics, effectively coarsening in the time domain. Discrete time and discrete state Markov State Model (MSM) is a method that has recently been developed to study microsecond or even millisecond protein-folding kinetics, by compiling many shorter (nanosecond-scale) atomistic simulations [11–18]. In this section, we will survey the multiscale modeling methods for proteins in both length and time domains.

### Coarse graining in the length domain

Protein dynamics is governed by an underlying rugged free energy landscape. The number of local minima in this landscape increases exponentially with the system size. Therefore, sufficient sampling of the protein free energy surface using atomistic MD simulations is normally limited by computing power. Coarse-grained modeling provides an efficient way to alleviate the sampling problem by grouping multiple atoms into a single site. Coarse-grained models not only reduce the degrees of freedom of the system, but also smooth the potential energy surface [1, 19]. Thus, these coarse-grained simulations can run orders of magnitude faster than atomistic simulations.

Coarse-grained simulations can be parameterized based on information from atomistic simulations, knowledge from structural databases or even data from thermodynamic experiments [1]. In these simulations, different resolutions are employed in series rather than in parallel. Development of coarse-grained potentials for proteins can be traced back to the 1970s. In pioneering work by Levitt in 1975 [20], each protein residue was represented by three points: two for the main-chain and one for the side-chain. Since then, many protein coarse-grained models have been developed [8–10]. For example, Voth and Izvekov [21] developed the multiscale coarse-graining (MS-CG) potential, where force field parameters are extracted from atomistic MD simulations using a force matching procedure. Force field parameters obtained from this procedure are tabulated, and thus not restricted to any analytical functions. More recently, the Wu group [22] obtained parameters from both atomistic simulations and a coil library of high-resolution X-ray protein structures for their PACE coarse-grained force field [22]. The PACE force field has been shown to successfully fold small proteins [23].

Coarse-grained simulations can make the computations much more efficient. However, the fine-grained degrees of freedom also play important roles in many biological processes, so it is difficult to derive a single coarse-grained representation for the entire system that is both economical and accurate. For this reason, many research groups have recently combined fine-grained and coarse-grained representations in a single mixed-resolution simulation [1]. The hybrid Quantum Mechanics and Molecular Mechanics (QM/MM) simulations are one example of mixed-resolution simulations, where two resolutions are modeled in a single simulation system [8]. In another study using MS-CG method, Shi et al. [24] mixed all-atom and coarse-grained
model to simulate an ion channel embedded in the lipid bilayer. In their model, the ion channel was modeled at all-atoms resolution, whereas the lipid and water were modeled at coarse-grained resolution. They showed that their mixed-resolution simulations reproduce well the atomistic simulations, and are significantly faster. The Schulten group [25] has recently simulated the complex between the lac repressor protein (LacI) and a DNA segment using a mixed resolution model. In their simulation, the protein was described using the all-atom model, while the DNA loop connecting the two protein operators was described using a mathematical model mimicking a continuous elastic ribbon. Using this model, they revealed that the rotation of the head group is essential for the function of the LacI and further identified key residues that may lock LacI in a particular configuration.

Models at different resolutions can also be mixed in the framework of the replica exchange method (REM). Lyman and Luo [26] have developed a dual-resolution REM, where two replicas are simulated at high and low resolutions simultaneously. Exchanges of configurations are attempted periodically between these two replicas so that the low-resolution replica can help overcome entropic barriers in the energy landscape of the high-resolution model. Lyman et al. [27] have developed an REM where high-resolution, as well as low-resolution models exchange to accelerate the sampling of the high-resolution models. One of the major challenges for these methods is how to efficiently reconstruct the high-resolution configurations from low-resolution ones. Recently, Christen and van Gunsteren [28] proposed the multigraining algorithm to help the reconstruction process by describing the coarse-grained particles as virtual particles in the atomistic model. Moreover, they have defined a grain-level parameter $\lambda$ to generate different replicas at intermediate levels of resolutions. They gradually change from atomistic model to the coarse-grained model; to keep a reasonable acceptance ratio, many intermediate levels are needed. More recently, Liu and Voth [19] proposed to relax the configuration at the coarse resolution before attempting the exchange. This ‘smart’ resolution REM scheme based on the smart walking method developed by Zhou and Berne [29] greatly increases the acceptance ratio so that only two replicas, one at the atomistic level and the other at the coarse-grained level are needed. This new method has been shown to quickly search the protein folded structure and also to approximately reproduce the Boltzmann distribution of the atomistic resolution model.

Coarse graining in the time domain

The biomolecular free energy landscape contains metastable free energy basins separated by free energy barriers. The presence of these metastable states in biomolecular dynamics has been suggested by various experiments. For example, using single-molecule FRET experiment, Zhuang et al. [30] observed four docked conformational states of distinct stabilities for a single hairpin ribozyme. In another relaxation dispersion NMR study, Mulder et al. [31] have identified two metastable states of the T4 lysozyme, and one state has around 2 kcal/mol higher free energy than the other one. If one can coarse grain conformational space into these metastable states, the fast motions within long-lived metastable states can then be further integrated out by coarse graining in the time domain. Discrete time and discrete state MSMs can automatically identify metastable states and calculate their equilibrium thermodynamics and kinetics. MSMs partition conformational space into a number of metastable states, and the resulting model has fast intrastate but slow interstate transitions. This separation of timescales can ensure that the model is Markovian at a discrete unit of $\Delta t$ in time if $\Delta t$ is longer than the fast intrastate relaxation time. Under this condition, the probability of a given state at time $t+\Delta t$ depends only on the state at time $t$. This allows MSMs built from short simulations to model long timescale dynamics. MSMs have been successfully applied to study protein conformational dynamics in a number of systems [11–18, 32].

MSMs are also inherently multiresolution due to the hierarchy of the free energy landscape. One can vary the resolution of an MSM by varying the degree of coarse graining in time as determined by the lag time. A short lag time will result in a high-resolution MSM with many metastable states. This high-resolution model will capture a large number of local free energy minima separated by small barriers. A long lag time will result in a low-resolution MSM with only a few states, and each of these states may contain multiple local free energy minima. Huang et al. [33] have developed the Super-level-set Hierarchical Clustering (SHC) algorithm that can
construct MSMs at multiple resolutions using hierarchical spectra clustering at different super-density levels. SHC is shown to be able to produce MSMs at different resolutions using different super-density level sets.

Different resolution models can also be mixed in a single MSM by integrating coarse-grain and atomistic subsystems. Kasson and Pande [34] have recently proposed the cross-graining algorithm to couple coarse-grained and atomistic simulations using MSMs. In this method, both the coarse-grained and atomistic subsystems are decomposed into discrete metastable states. Transition between a pair of states in the joint space is simply treated as the product of two subspace transitions by assuming that transitions in different subspaces are not correlated. This cross-graining method may provide a general way for simulating mixed-resolution systems such as membrane proteins.

In addition to the MSMs, there also exist other approaches of coarse-graining in the time domain. For example, Izaguirre et al. [35] have recently developed a new scheme to perform multiscale dynamics simulations using normal mode analysis. They first separate fast and slow modes. They then chose to propagate the dynamics only in the slow modes using a Langevin equation. This will allow as much as 500 times longer time step than that of atomistic simulations, and gain up to 200 times speed-up for protein simulations.

RNA MODELING

As explained, modeling of protein structure and dynamics has reached a degree of maturity, though considerable challenges remain. Computational modeling of RNA, on the other hand, has lagged behind proteins, due to the more recent appreciation of its importance, the greater experimental difficulty of crystallization, a smaller number of workers, the incompletely understood role of co-transcriptional [36] and hierarchical folding [37], and the theoretical issues of modeling its high charge, among other reasons. RNA workers have responded to the physicochemical differences between RNA and protein by developing computational tools that in many ways differ from those that work with proteins. In this section, we will discuss how dynamics and structure prediction are currently done in RNA. We then describe how these are being extended to greater time and length scales, with the objective of reaching the mesoscale, where the dynamics of molecules meets that of cells and even tissues.

We will encounter the recurring theme of multi-resolution modeling. Following [1] what we did for proteins, we divide such methods into serial and parallel schemes. Serial methods include those in which RNA structure is solved by sampling from precompiled databases followed by evaluation [38, 39], training a coarse grained force field on such databases, then computing coarse-grained dynamics [40, 41], and potentially returning to fine resolution in a final step [42]. Parallel methods include those in which different molecules or regions are treated at different levels of kinematic resolution [3, 43], those in which the forces and kinematics are treated at different resolutions, and those in which time is coarse grained with multiple metastable states explored simultaneously [33].

RNA dynamics

The importance of RNA has only grown over the years, as its pervasive role in gene regulation has come to light. Computational methods for computing RNA dynamics have encountered roadblocks, which are not as important in the world of protein dynamics. Part of the challenge is that the RNA is a large, highly charged, very flexible molecule with a dearth of the distinctive surface features needed for recognition [44] and a propensity for kinetic traps. These issues in particular, challenge the widely acknowledged gold standard method of protein motion calculations, molecular dynamics. In response, the RNA community has experimented with multiple methods that restrict fine-grained calculations to selected regions. For instance, the Q program treats a small spherical region within a system using conventional MD, and surrounds this sphere with a restrained layer of water molecules [45]. This is useful for calculations (e.g. Free Energy Perturbation or Linear Interaction Energy) in a small region within a larger complex; the error in free energy change is on the order of 1 kcal/mol [45]. In an alternate approach, some workers decrease the granularity of the forces for the entire system [5, 40, 41]. For example, in Ref. [5] structural statistics are used to train a potential acting on a subset of atoms; this is useful for discriminating near-native structures from poorer quality decoys. Coarse-grained forces are typically knowledge-based (KB) [5] rather than physics-based. However, not all KB potentials are coarse-grained; some of the most
successful are atomistic [5, 6]. KB force fields have the advantage that they may not only be faster, but also circumvent theoretical roadblocks that often challenge physics-based methods [5, 6]. As with forces, the granularity of kinematics can be reduced either globally, as in [40, 41] where the velocities are only updated when pseudoatoms enter or leave each other’s discrete neighborhoods, or in a region-dependent way [43]. Furthermore, the granularity of the kinematics can be the same as [40, 41] or different from [4] the granularity of the forces. The time integration methods are also more diverse, including variable time step [46] and collision-driven integrators [40, 47]. Last, it is possible to coarse-grain time and compute the probability of transitioning between metastable intermediates; for example in [33], many independent short-time trajectories of short RNA strand dynamics were connected to elucidate the statistical landscape of hairpin folding.

As mentioned, MD can be used for RNA modeling. It can even be used for large systems, famously including the ribosome [48]. However, the very long time scales require that such systems cannot be run unbiased until convergence [48]. One possibility for making unbiased physical simulation tractable is restricting the calculation to a sphere large enough to contain the region of interest, but not necessarily the entire solute molecule [45, 49]. Lastly, it is possible to effectively coarse-grain time using MSMs as described [33].

Many workers have abandoned physics-based force fields altogether in favor of KB force fields. However, these KB force fields have the disadvantage that having been trained on specific sets of observed phenomena, they cannot be expected to recapitulate phenomena not represented in training sets. Nevertheless, their saving grace is faster convergence, due to reduced degrees of freedom and/or force evaluations. One example of this approach is Discrete (DMD), which reduces complexity in three ways: (i) each residue is represented by three pseudoatoms, (ii) the potential is discontinuous, consisting of square wells and (iii) the time integration takes advantage of the resulting piecewise-constant particle velocities to update positions and velocities following a collision list [40, 47]. The Nucleic Acid Simulation Tool (NAST) reduces dimensionality even further, using a single pseudoatom to represent each residue; the user must provide base-pairing interactions to restrict the conformational search [41]. NAST can fold tRNA (within 8 Å RMSD of the native structure) and the P4P6 domain of the Tetrahymena Group I Intron (within 16 Å). MacroMoleculeBuilder (MMB) is more interactive; the force field consists of base-pairing interactions (of any type catalogued in [50]) and collision detecting spheres [46] (for preventing steric clashes), all specified by the user. Its internal coordinate framework [46] allows different regions of the molecule to be treated with different flexibility, e.g. any stretch of residues can be internally rigid for cost savings. It has been used to fold tRNA (within 9.6 Å) and P4P6 (within 10 Å) using biochemical and limited biophysical data [4]. It was also used to make a threaded model of a 200-nt ribozyme, coming within 4.6 Å of the native structure [3]. Last, it was able to easily model ribosomal translocation [43]. Due to its internal coordinate framework [51] it is particularly useful for modeling large complexes.

Thus, the state of the art in RNA dynamics has advanced significantly, with various methods found to reduce the problem space, degrees of freedom, forces and integration cost. These approaches have been useful for systems spanning a large size range, including the ribosome [43, 48, 49], ribozymes [3], tRNA [41] and small hairpins [33].

Structure prediction
Dynamics plays an important but not exclusive role in RNA structure prediction. There is no reported case of any RNA structure being solved by directly integrating the all-atoms equations of motion for the entire trajectory of folding, as has been done for proteins [52]. However, some structure prediction methods do use dynamics to minimize a coarse-grained potential, trained on structural data [40, 41]. These knowledge-based potentials can also be used at different resolutions to score the best high-resolution structures in an incremental way [5, 6]. In many cases, biochemical, biophysical and other specific nonstructural knowledge is used to restrain the problem [4, 41]. Other methods are not dynamical at all. A very popular approach is to assemble molecules using fragments drawn from structural databases, and evaluate these structures against a potential [39, 53]. Often, the structural sampling is done at a different resolution from the evaluation [39]. Fragment assembly is related to homology modeling [3, 54], which is becoming increasingly viable for RNA as the number of solved 3D structures increases.
The dynamical structure prediction codes include the mentioned DMD, NAST and MMB. DMD has successfully folded tRNA [40]. NAST [41] and MMB [4] have folded tRNA, as well as the P4P6 domain of the Tetrahymena Group I Introns. The latter two systems are the largest that have been solved using template-free dynamical methods. Threading can be done dynamically, as has been done for the entire Azoarcus Group I Intron [3].

Fragment assembly methods were introduced with MC-Sym [38], which has recapitulated numerous known structures [53] using a scoring function based on frequency of observation of fragments. Fragment Assembly of RNA (FARNA) similarly samples all-atom trinucleotide fragments from a structural database, and then evaluates the structures against a coarse-grained force field; it has predicted a number of small RNA structures de novo [55]. Like the dynamical methods, these have not yet been scaled to solve larger systems de novo. However, homology modeling can be done by fragment assembly, and this led to a model of the 16S subunit of the *Escherichia coli* ribosome [56].

**COMPLEXES, ASSEMBLIES AND AGGREGATES**

Having established protein- and RNA-specific modeling methods, we now go on to the analysis of large, possibly heterogeneous complexes. The function of a biomolecule largely depends on its interactions. Even if a large number of protein and nucleic acid structures are known, the structures of their assemblies remain mostly unknown. Modeling these assemblies is very complex as the number of degrees of freedom is large. Despite wide efforts and advanced techniques for studying protein and nucleic acid structures as described above, modeling assemblies and aggregates at different resolutions remains a challenge. Indeed upon interacting, the partners may undergo large conformational changes and the dynamics of such macromolecular machines are often intractable [57].

Docking is used to predict the structure of a complex when the individual structures of the components are known or can be modeled. Since the first description of a docking procedure in 1974, various techniques have been developed. They can be classified in two groups: rigid-body and flexible techniques. Their performances are evaluated in the community-wide experiment CAPRI (Critical Assessment of PRediction of Interactions) since 2001 [58]. It has shown interesting progress in the prediction of the interacting regions and in cases when flexibility is limited to small regions and changes [59].

**Multiscale docking and assemblies**

Protein–protein docking prediction techniques usually include three steps: finding putative complex conformations, scoring them to keep the most biologically relevant and refining the best scored structures [60, 61].

Finding suitable conformers involves 3-dimensional search and large conformational sampling. For very large assemblies, this cannot be easily achieved at the atomic level. Most of the protein–protein docking algorithms use coarse-grained representations for the initial sampling and scoring. To perform docking, the rigid-body procedures are widely used, however, it is also crucial to take flexibility into account for the partners.

Rigid body procedures imply finding putative candidates from the structures of the individual components, taken in their free (unbound) or complex (bound) form. They require an exhaustive spatial search for which many algorithms have been developed. Fast Fourier Transform (FFT) procedures are still commonly used [62–65], the level of representation being defined by the protein representation, such as those described for protein modeling (one, three or five pseudoatoms per residue mainly), or by the FFT grid size. The search is also done in direct space [66] or by geometric hashing [67], the protein models being coarse-grained. Aside from geometric hashing, these methods use algorithms that are computationally expensive and cannot deal with a fine grid size. The search is thus limited by the step size and may not lead to any usable results. The multiscale strategy involved is thus basic: reducing the grid/step size when external biological data or scoring functions provides information of putative epitopes. The strategy is successful in simple cases, for example protease-inhibitor complexes for which conformational changes are limited (e.g. a recent CAPRI example is the subtilisin Savinase—α-amylase subtilisin inhibitor BASI complex [59]).

Multilevel modeling is widely taken into account in scoring functions. A wide range of techniques are used, from data-driven docking, using conservation
or other experimental information [68], to machine learning techniques [69], physics-based force-fields [70] and statistical potentials [71]. Atomic details are often added after the scoring step to refine the prediction. The multilevel scoring step has shown to be a key part in the whole docking process [58]. As of today, scoring functions are still a bottleneck of docking procedures. They have yet proven able neither to predict binding affinities [72] nor to identify good conformations among a docking set in a blind setting [59].

Taking into account flexibility in a docking procedure is a very difficult task. Even if the flexibility is often limited to interface side chains [73], some complexes undergo large conformational changes as the docking benchmark shows [57]. Some docking procedures are able to deal with conformational changes and they make a great use of the different representation levels during the modeling. The RosettaDock suite is extremely well suited to this purpose. It can model backbone conformational changes using structural templates, model loops in free space and offer side-chain optimization either through a rotamer library or a well suited force field [74, 75]. Another approach is to use normal modes combined with different resolution levels to model the flexibility [76, 77]. In addition, geometric modeling has also been adapted [67, 78]. These methods mostly consider the molecules individually to model their flexibility and thus, cannot account for induced fit effects. They do provide some insights on the flexibility of the molecules but are often not accurate enough for describing the conformational changes involved in complex formation. For example, this sometimes leads to overly distorted models, such as the CAPRI results for various targets shows from Target 1 (HPR/HPR Kinase) [59, 79].

Multicomponent and symmetric docking can also be performed [67]. This is even a much harder problem but of great use when trying to fit experimentally obtained envelopes such as Electron Microscopy or Small Angle X-ray Scattering data. This problem is however, out of the scope of this review and will not be further discussed.

Due to its inherent complexity, protein–nucleic acid docking is lagging behind protein–protein docking. For example, some attempts have been made to predict RNA binding sites on proteins based on interaction propensity statistics combined with geometric calculations [80]. Some software suites such as PyDock [81] or HADDOCK [82] are also able to deal with protein–RNA and protein–DNA interaction prediction but benchmarking just recently appeared [83] and the CAPRI experiment conclusions show that the prediction techniques are not yet ready [59].

Despite a large number of published articles describing successful stories using multiscale procedures for various biological molecules of interest, automatic multiscale prediction with few or no biological external data is still limited. Over the recent years, the CAPRI experiment [59] and protein docking benchmark studies [57, 84] have shown that a satisfactory accuracy required for predicting interactomes [85] or binding affinity [72] has not yet been reached and results are still close to random when no or few external biological data is available. This may be due to the wide scale range the procedures have to accommodate, but also to the lack of efficient scoring functions both at the atomic and coarse-grained levels.

**Aggregation**

Protein aggregation and amyloid formation are key in the development of several diseases such as Alzheimer’s, Parkinson’s, Huntington’s and Creutzfeldt–Jakob’s. Computational modeling of protein aggregation has led to interesting insights on amyloid formation, such as the Ma-Nussinov-Tycko model for \( \alpha \)-amyloid [86].

All-atom studies of such systems are mainly based on techniques described above for proteins. They make a great use of MD simulations including different representations of the systems. Replica Exchange Molecular Dynamics has allowed large simulations and normal mode-based simulations can account for conformation changes and description of the most stable state. These methods are well reviewed elsewhere [86, 87].

Interestingly, specific coarse-grained models have been developed for aggregation modeling [88]. The level of coarse-graining ranges from a few beads per residue to a cuboid per oligomer. Therefore, studies using various resolutions of models lead to the description of the aggregation phenomenon at a wide length and timescale.

Many existing coarse-grained models for aggregation are residue-based. For example, in the PRIME model, each residue is represented using four spheres (three for the backbone and one for the side chain) and C\( \alpha \)-C\( \alpha \) distances are fixed using pseudo-bonds.
Simulations are made using Discontinuous Molecular Dynamics with hard sphere or square well potentials [89]. Another residue-based model proposed by Bellesia and Shea [90, 91], uses three points per residue and the backbone topology, adding a dihedral term to represent the flexibility and explicitly taking into account hydrogen-bonding and hydrophobic interactions for the peptide. This model is used in Langevin dynamic simulations. Another mid-resolution model used in Langevin dynamics simulations has been proposed by Caflisch and coworkers [92]. In this model, each peptide is described by four backbone beads and six side-chain beads. This model is not literally residue-based and partial charges are added to the backbone to represent the dipole. The flexibility is accounted for by a dihedral term populating a amyloid-protected state $\pi$ and an amyloid competent state $\beta$.

Models with a resolution lower than one bead per residue are mainly used in Monte-Carlo (MC) simulations. The lattice model is made of eight connected beads on a lattice. Beads are tagged hydrophobic or polar and the extremities of the lattice are charged [93, 94]. Simulations using the lattice model are performed with local and global move sets. The tube model is a lower resolution model [95] that is used in MC or Discontinuous Molecular Dynamics simulations. Each peptide is represented by a tube, not accounting for the residue sequence. The tube thickness indicates the volume exclusion and bending stiffness. Hydrophobic interactions and hydrogen-bonding effects are modeled through geometric constraints. The lowest resolution model used is the cuboid [96]. In this model, a cuboid can be used to represent an extended or folded peptide or a small oligomer called a building block. Conformational changes of a building block are ignored. The interactions made by the cuboid are described by three parameters, corresponding to the three pairs of opposite sides of the cuboid. The parameters describe strong attraction between cuboids in the intrasheet hydrogen-bonding direction, weak attraction in the intersheet direction and repulsion in the direction parallel to the cuboid building block. Only single-building block moves are considered in the MC simulations using this model.

Most of these models are used for $\alpha\beta$ amyloid formation studies. Despite the various coarse-grain sizes used, these models can often not be connected in a fully multiscale procedure. This may lead to different results at different scales and precludes yet, a full description strategy of $\alpha\beta$ amyloid formation, for example, from atomic resolution dimer formation to large fiber analysis.

**CONCLUSION**

We reviewed how novel multiresolution approaches are making inroads in structure and dynamics of protein, RNA and complexes. Many new special- and general-purpose force fields and potentials have been developed, with different force and energy granularities. Structures have been solved using a variety of dynamical, minimization and Monte-Carlo approaches, often with kinematic or sampling granularity that differs from that of the corresponding potential or force field. Similarly for the prediction of assemblies, many geometric representations are in use, while kinematics and potentials can change granularities from stage to stage in a calculation.

**Key Points**

- Knowledge-based potentials and force fields are available at different granularities for different purposes.
- Coarse graining can be done not only in the length but also in the time domain.
- RNA modeling has peculiarities that are treated with special-purpose force fields and potentials, again at multiple granularities.
- The modeling of assemblies is often achieved by shifting kinematic or force granularity, and treating special regions such as interfaces at different flexibility.

**FUNDING**

The Hong Kong Research Grants Council (grants 661011 and HKUST2/CRF/10 to X.H.); INRIA Equipe Associee Program (GNAPI) (to J.B); National Research Foundation of Korea (grant nos 305-20100005 and 2010-0001630 to S.S.); IBM Blue Gene Science Program; Swedish Research Council [the eSSENCE project grant (essenceof-science.se) to S.F.]; the Swedish Foundation for International Cooperation in Research and Higher Education (stint.se) (travel funding to S.F.).

**References**


