Discrete molecular dynamics

Elizabeth A. Proctor, Feng Ding and Nikolay V. Dokholyan*

Discrete molecular dynamics (DMD) has emerged as a simplification of traditional molecular dynamics (MD). DMD employs discrete step function potentials in place of the continuous potential used in traditional MD. As a result, the simulation engine solves the ballistic equations of motion for only those particles participating in a collision, instead of solving Newton's equations of motion for every particle in the system. Because fewer calculations are performed per time step, the DMD technique allows for longer time and length scales to become accessible in the simulation of large biomolecules. The use of coarse-grained models extends the computational advantage of this method. Although some accuracy is sacrificed to speed, because of the usefulness of DMD to the simulation of many particles at longer timescales, the technique has seen application to diverse molecular systems. © 2011 John Wiley & Sons, Ltd. *WIREs Comput Mol Sci* 2011 1 80–92 DOI: 10.1002/wcms.4

INTRODUCTION

The most problematic consideration in the field of molecular simulation is that of limited available computational power. Simplification of the system or of the performed calculations artificially increases the available computational power; less computer time is needed for each simulation step, leading to greater sampling of the studied system. Various simulation techniques such as molecular mechanics (MM) and molecular dynamics (MD), Monte Carlo, and Brownian dynamics, as well as hybrids of these methods, have emerged as simplifications of the exact quantum mechanical description for large molecules.[1] From these techniques, the most widely used simulation method in the field of molecular modeling is MM/MD. In MM, energies of the system in a certain conformation are calculated using a force field. This force field is a classical mechanical approximation of the full-quantum mechanical energy function for a system of particles. The energy function can be used as a criterion for local minimization, wherein a search algorithm is used to find an energetically optimal structure for the molecular system. In MD, in addition to energetics, molecular motions are also calculated by the iterative solution of Newton's equations for each particle simulated. Such simulations give an accurate energetic picture of a system and offer structural insight on an atomistic level. How-

ever, such accuracy is achieved at high computational cost for large molecules, including frequent system updates with small integration steps on the order of femtoseconds and electrostatic calculations scaling as the square of the number of particles, which limits the accessible timescale for the simulation of large molecules. As a result, a considerable amount of work has been put toward further simplifying the method. These efforts have been primarily put toward extending sampling techniques and reducing the system degrees of freedom, for example, by the development of implicit solvent models.

As a significant simplification of MD, two groups have proposed the use of discrete square-well potentials in simulations of biomolecules, a technique called discrete molecular dynamics (DMD).[2,3] This method employs two paths to simplification: the first through modification of the interaction potential and the second an innovative search algorithm. In DMD, the interactions between atoms are represented by the simplest possible potential function, a 'square-well' potential. This simplified method is a coarse-graining of the potential function itself and greatly reduces the number of calculations necessary, while sacrificing some of the accuracy found in a continuous potential function. In a stepwise potential, the derivatives of the potential are either zero for a constant-velocity regime or infinity when a collision occurs with an instantaneous change in velocity. Unlike MM and MD simulations, which are driven by physical forces, DMD simulations are driven by collision events as a consequence of the discretized nature of the interaction potential. The searching of these collision events gives rise to the second major

*Correspondence to: dokh@med.unc.edu

Department of Biochemistry and Biophysics, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

DOI: 10.1002/wcms.4

simplification present in DMD. Instead of searching all possible collisions, the search algorithm is limited to a local search; only those atoms within the same local area are considered as having possible collisions. Once collisions are identified, only the atoms participating in collision are required to be updated after each event, resulting in fewer calculations needed per time step and greatly increasing sampling ability. For example, in simulations of the 846-residue IgG3-Kappa antibody (PDB ID:1R24), an all-atom DMD simulation (see *Model Development* for a description of the all-atom protein model in DMD) can reach 360 picoseconds per day, whereas an implicit solvent MD simulation with no electrostatic force cutoffs will reach 19 picoseconds per day. These innovations make DMD simulations sufficiently fast that, with the use of coarse-grained models, the dynamics of large molecular complexes can be accessed on the timescale of seconds. Coarse-grained DMD can also be used in combination with all-atom MD to perform multiscale simulations, which are useful in revealing information about processes that occur in mid-length timescales such as the reconstruction of protein transition states.[4] At an adequately small step size, the discrete step function approaches the continuous potential function with an approximately linear decrease in the speed of simulations, and DMD simulations become equivalent to traditional MD.

In this article, we will discuss the usefulness of the DMD method to the simulation of molecular systems with many degrees of freedom. First, we will describe the DMD algorithm. Next, we will outline models of varying resolution that have been developed for distinct biomacromolecular systems. Finally, we will discuss applications of DMD to several systems of interest in current research.

ALGORITHM

The difference between DMD and traditional MD lies in their interaction potential functions. In approximating the continuous potential functions with step functions of pairwise distances, DMD simulations are reduced to event-driven MD simulations. Interatomic interactions in DMD are governed by square-well potential functions. Neighboring interactions describing the molecule geometry and polymer connectivity (such as bonds, bond angles, and dihedrals) are modeled by infinitely high-square-well potentials. The widths of these infinite square wells are determined empirically from the distribution of bond lengths between various atom types present in X-ray crystallographic structures. The assigned square-well width is the larger of the mean ± standard deviation
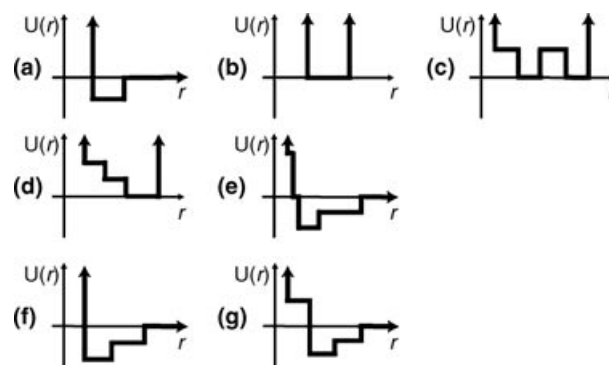

FIGURE 1 | Discrete molecular dynamics (DMD) potentials. Schematics of the various step function potentials used in DMD. (a) Hard-sphere interaction potential, characterized by infinite repulsion at the hard-sphere radius, attractive potential well, and zero interaction after a given radius. (b) Single-infinite square well used for covalent bonds, angular constraints, and base-stacking interactions. (c) Dihedral constraint potential. (d) Hydrogen-bonding auxiliary distance potential function. (e) Discretized van der Waals and solvation nonbonded interactions potential. (f) Base-pairing interaction potential. (g) Lysine–arginine-phosphate interaction potential in DNA–histone nucleosome complex.

of this distribution or the mean ± 2% of the mean value. The width of the covalent bond square well must be carefully chosen; a well that is too wide will cause deformations in molecular structure, whereas a well that is too narrow will create inefficiencies in the algorithm due to too frequent collisions.

The DMD interaction potentials are a set of spherically symmetric stepwise functions of the distance r between atoms. Atoms are described by their types, each type having characteristic mass and hard sphere and interaction radii. The pairwise potential energy is then assigned by a distance-determined stepwise function of level k, such that $r^{ab}_{min} < r^{ab}_1 < r^{ab}_2 ... < r^{ab}_k ... < r^{ab}_{max}$. If $r^{ab}_k < r^{ij} < r^{ab}_{k+1}$, then the pairwise potential energy is assigned as $U^{ij} = U^{ab}_k$ [see Figure 1(a) for a schematic of a single-step potential].

As implied by the descriptive term 'collision driven', all atoms in a DMD simulation move with constant velocity until they collide with another atom. In terms of the potentials discussed above, as soon as $r^{ij} = r^{ab}_k$, the potential of interaction between the two atoms changes, accordingly, the velocities of the two interacting atoms change instantaneously. These velocity changes are required to be in accordance with the conservation laws of energy, momentum, and angular momentum. This process will repeat as long as the atoms have sufficient kinetic energy to overcome the potential changes $\pm(U^{ab}_{k-1} - U^{ab}_k)$, where the sign depends on the forward or backward collision. Once

the kinetic energy is overbalanced by the potential change, the atoms undergo a hardcore elastic reflective collision with no change in potential energy. Each such collision is termed an 'event'.

Because only those atoms participating in the collision are required to update at each event, only one calculation is needed per event to determine the new state of the protein. These calculations are tracked in the collision table, which lists all possible collisions in the simulation. The amount of information contained in the collision table is further simplified by the use of a cubical grid. The simulation box, which contains all atoms used in the simulation, is divided into cubical cells. These cells have large enough dimensions to contain the interaction range between particles. Use of the cubical grid simplifies and reduces the search space for the determination of collisions by only considering and adding to the collision table the possible collisions within each cell.[5] The soonest collision is chosen from this table, between atoms i and j in time Δt, after which the collision table is updated with the new states of atoms i and j. All possible collision events listed in the table are then updated with the new velocities and positions of i and j, and any possible collisions with atoms that have moved out of range are deleted. New possible collisions are incorporated into the table, and the system time is advanced by Δt.

MODEL DEVELOPMENT

Models have been created to accommodate varying biomolecular systems. Depending on the complexity, flexibility, and desired resolution scale of the system of interest, molecules may be represented atom-by-atom or at a coarse-grained level. Simplified models are often desirable in that they speedup the simulation process and allow for access to greater timescales. However, when a greater level of detail is needed, an all-atom model is the best choice.

Protein Models

Throughout the years, proteins have been modeled to various levels of complexity, corresponding to the computational power available and time and length scales of interest. Because the majority of computational time in a DMD simulation is used in the calculation of bond oscillations, the computer time used to reach the same physical time roughly scales to the number of bonds in the protein system. Thus, the two-bead model will use approximately twice the computational time of the one-bead model, and the four-bead model will use approximately twice the computation time of the two-bead model. Be-

cause of the more complicated nature of the quasi-all-atom and all-atom models, computational time can as much as triple from the four-bead model to the quasi-all-atom model and double from the quasi-all-atom model to all-atom DMD.

Because of different levels of coarse-graining in the representation of residues, these models have different strengths to address the various problems of interest.[6] Coarse-grained models are useful to address properties of dynamics that are mainly governed by native structures of proteins instead of the amino acid sequences such as protein-folding kinetics,[7,8] domain swapping,[9,10] and the structural properties of denatured proteins.[11] Intermediate models are appropriate to answer dynamic problems wherein the accurate modeling of the backbone dihedrals and formation of backbone hydrogen bonds are required, such as the transition of secondary structures[12] and protein aggregation.[13,14] The quasi-all-atom model is useful in the pursuit of problems wherein details in both the backbone and the side chains of the residues are important, such as *ab initio* protein folding[15] and protein aggregation.[16] The all-atom model is the most sophisticated and finest-grained DMD model and is used to capture the atomic details of protein structure and dynamics.[17]

One-bead Model

In the one-bead model, each amino acid of the protein sequence is represented as a single sphere, with the entire protein represented by a 'beads-on-a-string' polymer model.[3] The DMD potentials make use of the structure-based Gō model[18,19]; those residues which are in contact in the native structure are assigned an attractive potential, and those that are not in contact are assigned a repulsive potential. A matrix of contacts is assigned so that, for residues i and j, if the native contact distance r_{ij} is such that the attractive spheres of the residues overlap, the element matrix Δ_{ij} has a value of 1. If r_{ij} is greater than this maximal interaction, then Δ_{ij} is assigned as -1. The contact matrix is then used to calculate pairwise interaction potentials, and the potential energy of the structure is a sum of these pairwise interactions, $E = \frac{1}{2} \sum_{i,j=1}^{N} U_{ij}$. This model has been used for protein-folding studies, wherein it was shown to admirably capture the thermodynamic and kinetic properties of folding dynamics.[3]

Two-bead Model

In the two-bead model, each amino acid is represented by one sphere for the C_α and one sphere for the C_β, with the exception of glycine, which has only the C_α
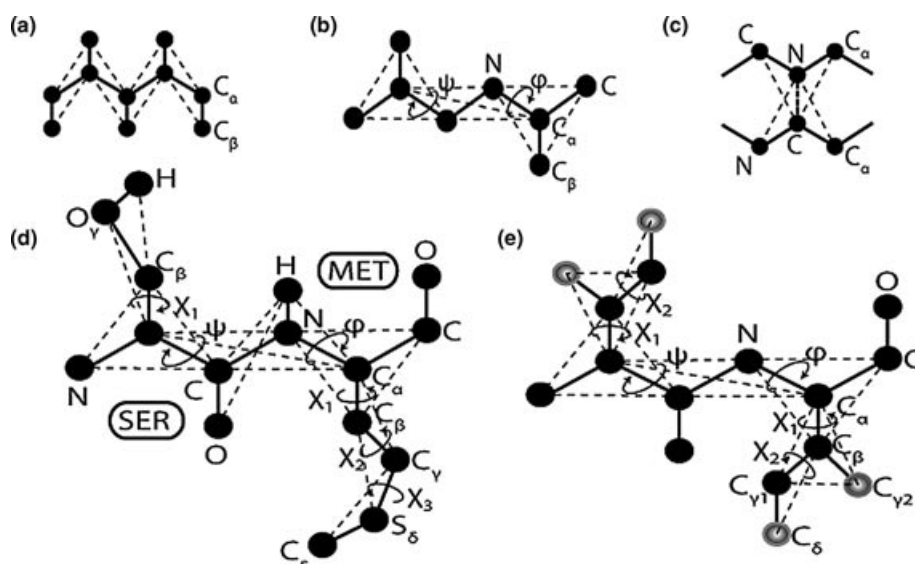
FIGURE 2 | Coarse-grained protein models. Schematics for coarse-grained models of varied resolution used for proteins in discrete molecular dynamics simulations. Circles represent beads or atoms, with gray circles being optional beads that may or may not be used for all amino acids. Thick, solid lines represent covalent bonds, whereas thin dashed lines represent angular constraints. Elliptical arrows represent dihedral constraints. (a) Two-bead protein model. (b) Four-bead protein model. (c) Hydrogen bonding with the reaction algorithm; thick dashed line represents the hydrogen bond. (d) All-atom protein model. (e) Quasi-all-atom protein model.

[see Figure 2(a)]. The introduction of an additional bead per residue limits the protein backbone flexibility, which in turn allows the modeling of high folding cooperativity and the characterization of protein transition states.[20] The transition-state ensemble is structurally interesting because of the ability of a protein in this collection of conformation states to change rapidly from folded to unfolded forms. This property is the focus of many protein design efforts.

Four-bead Model

In the four-bead model, each amino acid is represented by three backbone spheres (C, C_α, and N) and one sphere for the C_β, again with the exception of glycine [Figure 2(b)]. Additional backbone atoms are introduced for the modeling of N_i–C_j hydrogen bonds. Hydrogen-bonding interactions are modeled by the reaction algorithm.[12] Upon the formation of a hydrogen bond between nonadjacent residues i and j, N_i and C_j change their atom types to N'_i and C'_j, respectively. Whether or not a hydrogen bond will form depends not only on the interacting atoms but also on interactions between the neighbors of these atoms. To mimic the orientation-dependent hydrogen bond interaction, distances between neighboring atoms are included in the calculations, in addition to the distance-dependent interaction between N_i and C_j. Auxiliary interactions occur between atom pairs N'_i–$C_{\alpha j}$, N'_i–$N^{[\cdot]}_{j+1}$, C'_j–$C_{\alpha i}$, and C'_j–$C^{[\cdot]}_{i-1}$

[Figure 2(c)]. Once N and C reach the appropriate distance range for interaction, the auxiliary interaction distances, which define the orientations of the hydrogen bond, are evaluated to assign the total potential energy change between N_i–C_j and other surrounding atoms before and after the putative hydrogen bond formation. A two-step potential is used when assigning energy to auxiliary interactions to allow angular distortion with an energetic penalty [Figure 1(d)]. If the auxiliary interaction distances satisfy the predetermined range and the total kinetic energy of the interacting atoms is enough to overcome the potential energy change, hydrogen bond formation is allowed. If these conditions are not satisfied, a hydrogen bond does not form. The reaction algorithm for hydrogen bonding has been further adapted for the four-bead model by Nguyen et al.,[21] using a single-square-well potential instead of a two-step potential for simulating larger systems of protein aggregation.

The intermediate resolution of the four-bead model introduces a significant amount of extra calculation over the two-bead model, between the additional beads used in the model and the necessary hydrogen-bonding interactions introduced in the backbone. One method to compensate for these extra calculations is to introduce experimentally determined constraints to the protein system to decrease the accessible conformational space.[22] A set of inter-residue proximity constraints reduces the number of degrees of freedom of the system, and therefore the

number of possible collisions and amount of calculation are necessary.

This model was used in studies of transitions between different secondary structures,[12] wherein representation of all backbone atoms resulted in the ability to distinguish formation and transitions in secondary structure in peptides. The four-bead model has also been used in studies of helix bundle assembly,[23] wherein its intermediate resolution allowed for accurate modeling of backbone–backbone hydrogen bonds.

Quasi-all-atom Model
In the quasi-all-atom model, each amino acid is represented as in the four-bead model, with the addition of a C_γ atom to all residues except glycine and alanine. The beta-branched residues, threonine, isoleucine, and valine, have a second C_γ atom to represent the two braches after the C_β. The bulky residues, arginine, lysine, and tryptophan, are assigned an additional C_δ atom to simulate their additional volume [Figure 2(e)]. This model is useful for modeling protein core packing and side-chain degrees of freedom.[15] The quasi-all-atom regime improves on the four-bead model in that it can account for side-chain entropy and the effect of side-chain size on core packing. Borreguero et al.[24] further extended this model to include atoms other than carbons in a united-atom model. In this bridge between the quasi-all-atom model and the all-atom model, all protein atoms except hydrogens are explicitly represented.

All-atom Model
The all-atom model of DMD evolved from the need to study conformational dynamics on the atomic level. In the all-atom model, all heavy atoms and polar hydrogen atoms are explicitly represented.

In an all-atom model developed by Ding et al.,[17] bonded interactions are modeled using constraints on the covalent bond length, bond angles, and dihedral angles [Figure 2(d)]. For covalently bonded atoms i and j, interactions are modeled by a square-well potential [Figure 1(b)] such that the potential U^{ij} is equal to $+\infty$ for r^{ij} outside the range of the bond length $\pm$ the bond variance. This effectively renders covalent bonds as unbreakable, and two covalently bound atoms will be permanently constrained by this interaction for the duration of the simulation. Next nearest neighboring atoms i and j, with $j = i + 2$, interact under angular constraints represented by an infinitely high-square-well potential analogous to and dependent on that of the covalent bond constraints for consecutive atoms [Figure 1(b)]. Dihedral interactions between atoms i and j, with $j = i + 3$, are

modeled by multistep potential functions[15] of pairwise distance r^{ij}. The set of distance parameters (d_{min}, d_0, d_1, d_2, and d_{max}) for these potentials are experimentally determined from distance distributions in a nonredundant database of high-resolution protein structures [Figure 1(c)].

van der Waals (VDW) and solvation interactions are modeled as pairwise functions of distance. VDW interactions use the standard Lennard–Jones potential, and solvation interactions are modeled by the Lazaridis–Karplus solvation model.[25] Both functions are discretized by a multistep square-well function [see Figure 1(e)], characterized by the hardcore distance and a series of potential steps, similar to that described above for dihedral interactions.

Hydrogen-bonding interactions are modeled by the reaction algorithm,[12] which has been adapted to all-atom representation.[17] All possible interactions between backbone–backbone, backbone–side chain, and side chain–side chain atoms are included. Long-range electrostatic interactions are not included in the calculation of energies in the all-atom model due to the strong screening effect of solvent.

A similar all-atom model was developed by Emperador et al.[26] with a simplified nonbonded potential consisting of a two-step potential. A negative square-well potential is defined for hydrophobic interactions, and otherwise a finite barrier is used. To prevent unfolding artifacts caused by the simplicity of the nonbonded interaction potential, a compensatory square-well potential is defined between hydrogen-bonded residues, reinforcing secondary structural elements.

Luo et al.[27] developed an all-atom model based on the Gō model. Bonded interactions, covalent bond lengths and angular constraints, were represented as infinite square-well potentials. Nonbonded interactions are modeled with a two-step potential; at $r < 0.8r_{vdW}$, the potential is $+\infty$; within $0.8r_{vdW} < r < 1.2r_{vdW}$, the potential is attractive, and at $r > 1.2r_{vdW}$, the potential is zero.

Ribonucleic Acid Model
Because of the increased degrees of freedom in ribonucleic acid (RNA) as compared with proteins, RNA has a much greater flexibility, which creates a large challenge in the computational modeling of RNA. The conformational space of RNA increases exponentially with length, making traditional simulation techniques prohibitively computationally expensive. For this reason, a coarse-grained model of RNA is currently employed with DMD instead of an all-atom model.[28] RNA is modeled as 'beads-on-a-string', with
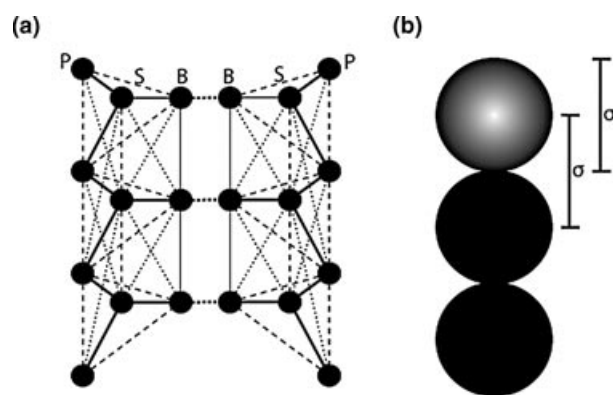
FIGURE 3 | Coarse-grained models for biomacromolecules. (a) RNA and DNA three-bead model. Black circles represent beads, thick solid lines represent covalent bonds, dashed lines represent angular constraints, thin dotted lines represent dihedral constraints, thin solid lines represent base-stacking interactions, and thick dotted lines represent base-pairing interactions. (b) Lipid three-bead model. Black beads represent fatty acid chains, whereas gray bead represents both head group and glycerol backbone. All beads have a radius of σ.

beads representing the sugar and phosphate groups, positioned at the respective centers of mass of each group, and the nucleobase, positioned at the geometric center of the hexagonal base ring [Figure 3(a)]. In a similar way as in proteins, neighboring beads are constrained by bond length, bond angle, and dihedrals. The parameters for these bonded interactions are calculated from high-resolution known RNA structures. Nonbonded interactions important for the determination of secondary and tertiary include base pairing, base stacking, phosphate–phosphate repulsion, and hydrophobic interactions. Parameters for these interactions are experimentally determined.

In combination with the coarse-grained nature of the RNA model, the rapid conformational sampling provided by the DMD algorithm allows for the efficient exploration of RNA conformational space, with computational time scaling linearly with length. The number of degrees of freedom in the RNA system can be further limited using experimental constraints such as sets of pairwise internucleotide distances.[29] These three techniques used in concert sufficiently simplify the RNA system to a degree where computational simulation is feasible and practical.

Nucleosomal Deoxyribonucleic Acid Model

In a similar way to RNA, DMD has been used to simulate nucleosomal deoxyribonucleic acid (DNA) in complex with histone using a 'beads-on-a-string' coarse-grained model for DNA.[30] The beads are the same as those used in the RNA model described

above: one bead for sugar, located at the centroid of the deoxyribose ring; one bead for phosphate, located at the centroid of that group's constituent atoms; and one bead for the base, located at the centroid of the hexagonal ring [Figure 3(a)].

Bonded interactions, consisting of covalent bonds, bond angles, and dihedrals, are modeled as infinite square-well potentials [Figure 1(b)]. Base-stacking constraints are also modeled by an infinitely high-potential well [Figure 1(b)], whereas sugar-base steric constraints are defined by hard-sphere repulsion. Base-pairing interactions are modeled as two-step potential functions [Figure 1(f)]. Interactions between DNA and histones consist of hydrogen bonding between the basic residues lysine and arginine of the histone and the acidic phosphates in DNA. These interactions are modeled as three-step potential functions [Figure 1(g)]. Because of the negative charge of the DNA molecule, interactions between DNA and the acidic or nonpolar residues of histone are negligible, and are not included in the model.

Lipid Model

A three-bead DMD lipid model has been introduced to study the self-organization of lipids.[31] The first bead represents both the lipid head group and the glycerol backbone, and the second and third beads each represent a fatty acid chain [Figure 3(b)]. All beads are of identical diameter, σ, which is also equal to the bond lengths. Both bonding and flexibility potentials are represented as discrete square-well functions. Flexibility is governed by the angle formed between the two bonds in the three-bead chain. The bond angle potential is also represented as an infinite square well as a function of distance between the first bead and the last bead.

Intermolecular attractive interactions between fatty acid chains are modeled by a two-step discrete potential. For distances between σ and $\sigma + w$, where w is an adjustable parameter, the potential is $-\varepsilon$. The tail–tail potential is set to zero for distances greater than $\sigma + w$ and infinity for less than the hard-sphere radius. Head–head and head–tail intermolecular pairs were defined as noninteracting, with the exception of hard-sphere repulsion.

APPLICATIONS

The power of the DMD engine has been harnessed to study various systems applicable to biology. The advantages of DMD are the most apparent in systems with many degrees of freedom, such as biological macromolecules and explicit solvents. Utilizing

DMD methods, these systems may be explored to biologically relevant timescales.

Protein Misfolding and Aggregation

Protein aggregation has been implicated in numerous neurodegenerative diseases, including amyotrophic lateral sclerosis (ALS, also known as Lou Gehrig's disease) and Alzheimer's disease. The long time and length scales relevant to protein misfolding and aggregation make DMD a valuable tool for use in probing the biological mechanisms of these phenomena, especially when used in combination with simplified protein models such as those described above.

Biologically, there are many paths that a protein may take to aggregation. Researchers study these scenarios in proteins or domains that are known to aggregate in order to understand the mechanisms behind the pathogenic association of proteins. Often, aggregation starts because of a misfolding in the protein, exposing hydrophobic core areas that then form favorable associations. In DMD simulations of the protein Cu, Zn superoxide dismutase (SOD1), implicated in ALS, Khare et al.[32] identified residues important to the folding kinetics and postulated that mutations to these residues disturb important interactions within the protein, ultimately leading to misfolding and then aggregation. Other pathways to aggregation include helix bundle assembly[23] and conformational changes,[33] which have been simulated using DMD and four-bead protein models with good agreement to experimental results.

Because of the large time and length scales involved in the aggregation of proteins, as well as the large number of particles, the simplified potentials and fewer calculations utilized in DMD, along with coarse-grained models, are valuable tools for exploring the dynamics of aggregate formation. Ding et al.[9] used DMD along with a simplified protein model to predict the formation of amyloid structures in Src SH3 domains. The authors find two aggregate conformations; one of those could be capable of forming large amyloids. Further understanding of exactly how these aggregates form could elucidate a general mechanism for the formation of amyloids in larger, disease-relevant proteins.

Protein Folding

One of the oldest questions in the study of biophysics is that of the protein-folding problem: how is the three-dimensional structure of a protein determined by its sequence? An understanding of how proteins fold into their respective structures is necessary for identifying the link between protein structure and function and to be able to manipulate a structure to produce a desired function. Because of the vast complexity of this problem, a common approach to understand folding kinetics is from the study of protein crystal structures.[34] Crystal structures provide valuable information about the interactions between residues, and the computational study of these structures has resulted in many useful models. However, the folding timescale for most proteins is in the range of 1 ms to 1 s, which is prohibitively long for traditional atomistic MD simulations. Coarse-grained DMD, with its fewer and more efficient calculations, can reach longer timescales, and is an excellent tool for the study of protein-folding mechanisms.

Despite the interest in and tenure of the protein-folding problem, exact mechanisms remain a mystery, and little success has been achieved with large or even medium-sized proteins, mainly because of the many degrees of freedom present in the completely extended state.[35] In exploring this problem, researchers, therefore, often explore the folding mechanisms of domains or very small proteins to further reduce the number of calculations necessary. This is especially true when folding the protein *ab initio*, starting from the unfolded or fully extended state. Coarse-grained models also reduce the conformational space, although some possibly important interactions can be lost. To this purpose, Ding et al.[15] developed a quasi-all-atom protein model (see *Model Development*) to include consideration for protein backbone flexibility and side-chain packing while retaining some of the benefits of coarse-graining. The authors utilize DMD to conduct *ab initio* folding simulations of the 20-residue protein Trp-cage.[15] The model consistently folds from the fully extended unfolded state to a near-native set of conformations, with the backbone within 2 Å root mean square deviation (RMSD) of that of the nuclear magnetic resonance (NMR)-determined structure. Some observed states reached a backbone conformation with less than 1 Å of the average NMR structure.

For greater accuracy, although more computationally expensive, an all-atom model (see *Model Development*) can be employed in similar situations. Ding et al.[17] performed DMD folding simulations of six distinct small proteins (20–60 residues). To compensate for the shorter timescale available when using an atomistic protein model, the authors utilized the replica exchange method,[36,37] running parallel simulations at various temperatures, swapping them at time intervals according to a Monte Carlo probability. Such exchange facilitates uniform sampling of conformational space by overcoming energy

barriers while maintaining the free energy surface of the protein. The authors performed simulations using both the solvent-dependent reaction algorithm for hydrogen bonding or environment-independent hydrogen bonding. In the case of the reaction algorithm, all structures folded to native or near-native conformations. With environment-independent hydrogen bonding, either none of the proteins folded to distinct structures, and no secondary structure was formed, or incorrect secondary structures were formed. This study shows clearly that although the systems studied were very simple, environment-dependent hydrogen bonding is important to protein structure and folding.

Although *ab initio* folding simulations are invaluable because they allow for the folding of sequences whose structure is not already known, they are computationally expensive because the conformational space of a fully extended protein is so large. Another approach to the protein-folding problem, used to illuminate folding mechanisms when the native structure of a protein is already known, reduces this conformational space by biasing the conformational search toward the native state. This approach, called the Gō model,[18,19] assigns an energetic penalty to nonnative contact formations. Luo et al.[27] utilized all-atom Gō model DMD simulations to study the folding mechanisms of multiple beta-hairpins within a protein. The authors performed simulations on the Pin1 WW domain, which consists of two adjacent beta-hairpins. Two previously proposed models[38–40] of hairpin formation were both confirmed, with the folding following the one model at low temperatures and the other at the transition-state temperature. Which path the folding followed was found to be entirely temperature dependent, with no reliance on which hairpin formed first.

Protein Structure and Dynamics

At the heart of modern protein science is the study of protein structure and its relation to protein function. The structure–function relationship is best explored through the study of individual protein dynamics, which is difficult to examine with experiments alone. Simulation is a valuable tool because it allows for the three-dimensional analysis of a protein structure and the study of movements of individual side chains. Both traditional MD and DMD are used extensively in the study of protein dynamics. Which protocol to use is a question of system size, relevant timescale, and appropriate resolution. Emperador et al.[41] conducted a study comparing the usefulness of the DMD method with that of traditional MD for the description of

equilibrium protein dynamics. The protein model for MD was fully atomistic, whereas the authors utilized a one-bead protein model (see *Model Development*) for DMD simulations. It was shown that the one-bead model in DMD for most structures tested gives a reasonable reproduction of results found using all-atom MD with explicit solvent, and that such a coarse-grained model is especially useful in large proteins, wherein the most important dynamics are the larger movements of domain–domain interactions or loop fluctuations. All-atom detail is found to be necessary in cases in which large local nonharmonic changes take place.

DMD is a useful tool in the study of protein dynamics due to its ability to access longer timescales, and therefore reveals features that may not be otherwise observed. One application of DMD that is particularly useful in this regard is protein stability. Jang et al.[42] utilized DMD simulations to study the thermodynamic stability of a homotetrameric beta-sheet complex using a hybrid Gō-type potential. Simulations revealed three oligomer and four dissociated states. The state was dependent on both the temperature and the intermolecular contact parameter η, which measures the ratio of intermolecular to intramolecular native contacts. Emperador et al.[43] developed an all-atom model for DMD, utilizing a hybrid force field representing both physical potentials and the Gō model. In their model, a reference structure is used to define the Hamiltonian, but only those connections that are physically meaningful are used as Gō constraints. In this way, sampling of the structure is thorough, but unrealistic conformations are not represented. This model was then used to study protein flexibility and was able to accurately reproduce protein dynamics in solution.[26]

By understanding the structural parameters that keep a protein in the folded and/or associated state, a more complete link may be made between protein structure and function. This approach is especially important to disease etiology when the malfunction or destabilization of a protein contributes to human disease, and an understanding could lead to treatment or a cure. Ding and Dokholyan[44] used all-atom replica-exchange DMD to study the dynamic effects of demetallation and disulfide bond formation in SOD1, a protein implicated in ALS. Each SOD1 monomer binds one Cu and one Zn atom, as well as has a disulfide bond between Cys-57 and Cys-146. Metal coordination was found to greatly stabilize both the individual monomers and the dimer interface by minimizing structural fluctuations in the beta-strands and the interface. The disulfide bond also contributes greatly to dimer stability by anchoring one of the long loops,

although it has only a small effect on monomer stability. These structural insights into the dynamics and stability of SOD1 suggest mechanisms for disease in mutations affecting the areas near the metal binding and disulfide bond sites.

Protein Design

Protein design is a topic of high interest in many fields of science due to its broad possibilities for application, from industrial cleanup to pharmaceuticals and many other areas. The mirror image of the protein-folding problem, protein design, has historically presented a challenge because of the lack of perfect knowledge of how and why atomic interactions are formed and broken. Lack of the necessary computational power to reach protein-folding timescales has also contributed to difficulties, forcing the need for simplified and less accurate models of atomic interactions. To better understand the atomic interactions necessary for folding and to design stable proteins, three questions must be addressed[45]: (1) What is the alphabet of protein sequence? (2) How much flexibility is allowed in the protein model? (3) How may protein conformations be sampled without structural bias? The efficient sampling method of DMD can be applied as an answer to this third question.

A common design practice is to make mutations to an existing structure, using it as a scaffold for a new protein. This method uses the principles of protein evolution to create homologous (>30% sequence identity) or analogous (<20% sequence identity) proteins. One or two such mutations do not significantly affect the overall fold of the protein, but the accumulation of mutations leads to distinct packing in the protein core, which can lead to the creation of homologues and, if enough differentiation occurs, analogues.[46] This is the postulate tested by Ding and Dokholyan,[46] using a combination of approaches toward exploring sequence and structural conformation spaces. The authors utilized DMD to introduce structural perturbations to a scaffold, using energetics as a guide to search the conformational space of the reference structure to create a diverse ensemble of similar structures. The authors developed a molecular design and modeling suite, called Medusa, to then assign the most thermodynamically stable sequence for each structure in the ensemble. It was found that, as observed experimentally, core residues are conserved preferentially over surface residues up to the threshold of 20–30% sequence identity, which corresponds to the 'gray area' between homologous and analogous proteins. As the sequence further diversifies, core residues diverge as much as surface residues.

This finding is also consistent with the current theory of protein evolution.

Ribonucleic Acid

RNA molecules are of great interest to life scientists due to their important role in the central dogma of molecular biology, from its role in transcription and translation to ribosomal structure. In recent years, RNA molecules have been found to have much more diverse functions than only those described by the central dogma and are in fact ubiquitous to cellular processes.[47–49] For these reasons, more information about the structure and folding of RNA is greatly desirable, and computational simulation can provide an individual molecule focus impossible in experiments. The barrier to detailed simulations of RNA lies in the extreme flexibility of the nucleic acid molecule. The increased degrees of freedom over other macromolecules such as proteins, discussed earlier, cause exponentially more calculations to be necessary. For this reason, DMD is an important tool for the study of RNA structure due to its sampling efficiency and speed.

It is useful in a flexible system such as RNA to decrease the number of conformational degrees of freedom using a coarse-grained model. Ding et al.[28] developed a simplified model of RNA for use in DMD simulations (*Model Development*) and tested this model on the *ab initio* prediction of RNA tertiary structure and folding dynamics. The authors conducted replica-exchange folding simulations of 153 mid-length (10–100 nucleotides) RNA sequences whose structures have been experimentally determined; shorter sequences lack tertiary structure and so were excluded from the study. The simulations were initiated with the RNA sequence in its fully extended conformation, and the lowest energy conformation was taken as the predicted structure. For the majority of sequences, the predicted structures have Q-values close to 1, with an average value across all sequences of 0.94. For sequences less than 50 nucleotides in length, RMSD of predicted structures from the experimental structure was found to be less than 6 Å. For all 153 sequences, 129 had RMSD values of less than 4 Å. Such close correlation with crystal structures justifies the use of coarse-graining as an acceptable solution to the flexibility of the RNA molecule.

In addition to a coarse-grained model, experimentally determined constraints may also be imposed on DMD simulations to further reduce the conformational space. Gherghe et al.[29] used this method for the *de novo* determination of complex tertiary folds

by first determining experimentally the secondary structure and long-range inter-residue distances and applying this information to DMD simulations to construct a three-dimensional structure. Secondary structure was determined using selective 2′-hydroxyl acylation analyzed and primer extension chemistry,[50,51] whereas inter-residue distances were probed by placing a small moiety at various sites selective to motif. These data are used as input to the DMD engine, where base pairs form and weights are applied to the potentials according to the experimentally determined long-range interaction constraints. After a system equilibration, representative structures are collected from the simulations at evenly spaced time points. After selecting for low-energy, time-separated models, models were clustered and the final structure was selected as the central structure in the most populated cluster. Using the hybrid experimental and computational approach, the three-dimensional structure prediction of a test case of tRNA[Asp] was found to be within the range of 3.8–5 Å of the crystallographic structure, having as an input only the RNA sequence.

Chromatin

Unlike RNA, DNA in humans is not free in the cell, but is instead confined to the crowded nucleus, wrapped around histone proteins, and tightly packed into fibers of chromatin. It is still unknown how nucleosomal DNA is organized and packed into chromatin, although this process is vital to replication, transcription, and DNA repair. In a similar way to RNA, DNA molecules have a much higher flexibility, and therefore have much larger conformational space than proteins. An added difficulty not present in the simulation of RNA is the extremely large size of the nucleosome core particle (over 16,850 heavy atoms). These circumstances make the use of DMD for simulation of nucleosome core particles extremely useful to the study of dynamics at relevant time and length scales. As in the applications discussed above, the use of a coarse-grained model will reduce the number of particles in the simulation, and therefore decrease the amount of calculation necessary, increasing sampling efficiency. Sharma et al.[30] developed a simplified model for nucleosomal core particles, consisting of the histone octamer represented by a two-bead model, and each nucleotide of the 147 base pairs DNA fragment simulated by a three-bead model (see *Model Development*). This model was shown to successfully reproduce the stability and dynamics of a single nucleosome on experimentally relevant timescales. Histone tails were found to form salt bridges with associated DNA, suggesting that they play a significant role in overall chromatin structure. Histone tails were also shown to dominate nucleosome dynamics, with normal modes involving interhistone fluctuations.

Simulation of Liquids

The usefulness of the DMD technique is not only apparent in macromolecule systems but also in all systems with large degrees of freedom. Other examples of such systems can be found in liquids. In the biological sciences, the interactions between polar water and apolar solutes are important for the folding of proteins, formation of cell membranes, and many other important processes. This 'hydrophobic hydration' is especially important to the protein-folding problem discussed above, and a large body of theoretical and computational work exists discussing this phenomenon.

Water is extremely important to biological systems; but because of the large number of water molecules involved in solvation, simulations with explicit solvent can be very computationally expensive, given the special thermodynamic properties of water. Therefore, there is a need for a model of water that, although still explicit, is simplified. Buldyrev et al.[52] utilized DMD simulations to study the fluid dynamics of Jagla solvents; water-like anomalies described by spherically symmetric potentials with two characteristic parameters. These parameters are the hard core and soft core distances, the ratio of which can be carefully controlled. Such a model is called a ramp solvent. Within a very narrow range of this ratio, the Jagla model exhibits water-like structure and thermodynamics in a simple system. The authors find that when the ramp solvent is supplemented with an attractive tail, relative to apolar solutes, evidence of water-like hydrophobic hydration is exhibited. Specifically, solubility minima with respect to temperature and low-temperature swelling were observed. Expansion upon cooling is a common factor between water and the Jagla solvent; and although they occur by different mechanisms, this phenomenon suggests similar hydration dynamics, making study of the simpler Jagla fluids relevant to water dynamics. Such models for water could be implemented in folding and dynamics simulations to simplify solvation dynamics calculations.

Artificial Materials

The efficient calculation and sampling present in the DMD algorithm are also useful in engineering applications with artificially manufactured systems. Nanoparticles are of great industrial interest because

they are routinely used in the manufacture of technological products, as well as present in many consumer products, from photovoltaic cells to sunscreen. Simulation of nanoparticles reveals behaviors and properties that are difficult to observe in experiments. Schultz et al.[53] utilized DMD to study the phase behavior of copolymer–nanoparticle composite materials. In composite materials, copolymers serve as an ordered matrix for the nanoparticles, which improve the polymer strength or contribute to thermal properties. In this study, the copolymer was modeled as chains of 20 hard spheres, with a square-well repulsion potential between copolymer segments, and the nanoparticle was modeled as a sphere with various radii. Nanoparticles either have neutral interactions with both copolymers or repulsive interactions with one of the copolymer segments. Phase diagrams constructed according to nanoparticle radius and strength of nanoparticle repulsion from copolymer were found to be in agreement with previous Monte Carlo simulation results.[54] Concentration profiles and periodic spacing were also measured, showing that larger particles tend to show an increase in localization and have a decrease in affect on polymer concentration profile, and neutral particles decrease periodic spacing, whereas repulsive particles tend to increase periodic spacing.

Another engineering application of interest to industry is the construction of dense monolayers used in a variety of fields such as electrochemistry, spectroscopy, and catalysis. Self-assembled monolayers are invaluable to these applications, but are limited by their low density, allowing the penetration of solvents into the monolayer. A solution to this is the use of mechanically assembled monolayers (MAMs), which are molecular assemblies, whose surface density has been artificially increased beyond what would be found in a self-assembled monolayer. Strickland et al.[55] conducted DMD simulations of MAMs using a 'beads-on-a-string' model for the polymers. In the model, a number of polymers were tethered to a flat surface, and then this surface was periodically artificially compressed by uniformly decreasing the surface area. The height of the polymers from the surface as well as monomer density and end-monomer density closely match theoretical predictions as well as previous sets of simulations. It was found that use of a longer polymer chain resulted in complete surface coverage at less compression. These simulations provide valuable information to experimentalists for use in the efficient creation of MAMs.

CONCLUSION

DMD is a powerful and fast computational method used by an increasing number of people in both computational and experimental fields. The ability of DMD to reach large time and length scales is invaluable in the study of complex systems. Its speed of calculation and efficient sampling of conformational space are useful in applications across many disciplines, giving important structural and dynamical insight into experimental results and allowing the generation of experimentally testable hypotheses.[56]

The increased speed and sampling ability of DMD are due to a simplification and coarse-graining of the continuous potentials present in traditional MD. Speed and sampling are increased at the cost of energetic accuracy. This approach is advantageous for large systems in which high resolution is not necessary. In systems in which exact atomistic accuracy is needed, or that have energetic irregularities that would not be captured by discrete potentials, traditional MD is a better choice for simulations.

Although DMD has seen applications in biomolecules primarily to protein systems, it has also seen exciting development in other areas such as nucleic acids and nanoparticles. There is room for expansion into other topics, such as protein aggregates and multicomplex systems, as well as DNA–protein or RNA–protein dynamics and *trans*-membrane protein systems with the combined use of DMD and coarse-grained models. As available computational power increases, all-atom DMD may be applied to larger systems, but in the present time, coarse-grained models are the most promising tools for the large-scale exploration of many-particle systems at biologically relevant timescales.

REFERENCES

1. Allen MP, Tildesley DJ. *Computer Simulations of Liquids*. Oxford: Oxford University Press; 1987.

2. Zhou Y, Karplus M. Folding thermodynamics of a model three-helix-bundle protein. *Proc Natl Acad Sci U S A* 1997, 94:14429–14432.

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

4. Ding F, Guo W, Dokholyan NV, Shakhnovich EI, Shea JE. Reconstruction of the src-SH3 protein domain

transition state ensemble using multiscale molecular dynamics simulations. *J Mol Biol* 2005, 350:1035–1050.

5. Rapaport DC. *The Art of Molecular Dynamics Simulation*. 2nd ed. Cambridge: Cambridge University Press; 2004.

6. Dokholyan NV. Studies of folding and misfolding using simplified models. *Curr Opin Struct Biol* 2006, 16:79–85.

7. Dokholyan NV, Buldyrev SV, Stanley HE, Shakhnovich EI. Identifying the protein folding nucleus using molecular dynamics. *J Mol Biol* 2000, 296:1183–1188.

8. Borreguero JM, Dokholyan NV, Buldyrev SV, Shakhnovich EI, Stanley HE. Thermodynamics and folding kinetics analysis of the SH3 domain form discrete molecular dynamics. *J Mol Biol* 2002, 318:863–876.

9. Ding F, Dokholyan NV, Buldyrev SV, Stanley HE, Shakhnovich EI. Molecular dynamics simulation of the SH3 domain aggregation suggests a generic amyloidogenesis mechanism. *J Mol Biol* 2002, 324:851–857.

10. Ding F, Prutzman KC, Campbell SL, Dokholyan NV. Topological determinants of protein domain swapping. *Structure* 2006, 14:5–14.

11. Ding F, Jha RK, Dokholyan NV. Scaling behavior and structure of denatured proteins. *Structure* 2005, 13:1047–1054.

12. Ding F, Borreguero JM, Buldyrey SV, Stanley HE, Dokholyan NV. Mechanism for the alpha-helix to beta-hairpin transition. *Proteins* 2003, 53:220–228.

13. Urbanc B, Cruz L, Ding F, Sammond D, Khare S, Buldyrev SV, Stanley HE, Dokholyan NV. Molecular dynamics simulation of amyloid beta dimer formation. *Biophys J* 2004, 87:2310–2321.

14. Urbanc B, Cruz L, Yun S, Buldyrev SV, Bitan G, Teplow DB, Stanley HE. In silico study of amyloid beta-protein folding and oligomerization. *Proc Natl Acad Sci U S A* 2004, 101:17345–17350.

15. Ding F, Buldyrev SV, Dokholyan NV. Folding Trp-cage to NMR resolution native structure using a coarse-grained protein model. *Biophys J* 2005, 88:147–155.

16. Ding F, LaRocque JJ, Dokholyan NV. Direct observation of protein folding, aggregation, and a prion-like conformational conversion. *J Biol Chem* 2005, 280:40235–40240.

17. Ding F, Tsao D, Nie H, Dokholyan NV. Ab initio folding of proteins with all-atom discrete molecular dynamics. *Structure* 2008, 16:1010–1018.

18. 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.

19. Go N. Theoretical studies of protein folding. *Annu Rev Biophys Bioeng* 1983, 12:183–210.

20. Ding F, Dokholyan NV, Buldyrev SV, Stanley HE, Shakhnovich EI. Direct molecular dynamics observation of protein folding transition state ensemble. *Biophys J* 2002, 83:3525–3532.

21. Nguyen HD, Marchut AJ, Hall CK. Solvent effects on the conformational transition of a model polyalanine peptide. *Protein Sci* 2004, 13:2909–2924.

22. Chen Y, Ding F, Dokholyan NV. Fidelity of the protein structure reconstruction from inter-residue proximity constraints. *J Phys Chem B* 2007, 111:7432–7438.

23. Smith AV, Hall CK. Assembly of a tetrameric alpha-helical bundle: computer simulations on an intermediate-resolution protein model. *Proteins* 2001, 44:376–391.

24. Borreguero JM, Urbanc B, Lazo ND, Buldyrev SV, Teplow DB, Stanley HE. Folding events in the 21–30 region of amyloid beta-protein (Abeta) studied in silico. *Proc Natl Acad Sci U S A* 2005, 102:6015–6020.

25. Lazaridis T, Karplus M. Effective energy function for proteins in solution. *Proteins* 1999, 35:133–152.

26. Emperador A, Meyer T, Orozco M. Protein flexibility from discrete molecular dynamics simulations using quasi-physical potentials. *Proteins* 2010, 78:83–94.

27. Luo Z, Ding J, Zhou Y. Folding mechanisms of individual beta-hairpins in a Go model of Pin1 WW domain by all-atom molecular dynamics simulations. *J Chem Phys* 2008, 128:225103.

28. Ding F, Sharma S, Chalasani P, Demidov VV, Broude NE, Dokholyan NV. Ab initio RNA folding by discrete molecular dynamics: from structure prediction to folding mechanisms. *RNA* 2008, 14:1164–1173.

29. Gherghe CM, Leonard CW, Ding F, Dokholyan NV, Weeks KM. Native-like RNA tertiary structures using a sequence-encoded cleavage agent and refinement by discrete molecular dynamics. *J Am Chem Soc* 2009, 131:2541–2546.

30. Sharma S, Ding F, Dokholyan NV. Multiscale modeling of nucleosome dynamics. *Biophys J* 2007, 92:1457–1470.

31. Davis CH, Nie H, Dokholyan NV. Insights into thermophilic archaebacterial membrane stability from simplified models of lipid membranes. *Phys Rev E Stat Nonlin Soft Matter Phys* 2007, 75(5 Pt 1):051921.

32. Khare SD, Ding F, Dokholyan NV. Folding of Cu, Zn superoxide dismutase and familial amyotrophic lateral sclerosis. *J Mol Biol* 2003, 334:515–525.

33. Lam A, Urbanc B, Borreguero JM, Lazo ND, Teplow DB, Stanley HE. Discrete Molecular Dynamics Study of Alzheimer Amyloid beta-protein (Abeta) Folding. *Proc of the 2006 Intl. Conf. on Bioinf. and Comp. Biol.* 2006, 322–328.

34. Dokholyan NV, Borreguero JM, Buldyrev SV, Ding F, Stanley HE, Shakhnovich EI. Identifying importance of amino acids for protein folding from crystal structures. *Methods Enzymol* 2003, 374:616–638.

35. Grosberg AYu, Khoklov AR. *Giant Molecules: Here, There, and Everywhere*. San Diego: Academic Press; 1997.

36. Okamoto Y. Generalized-ensemble algorithms: enhanced sampling techniques for Monte Carlo and molecular dynamics simulations. *J Mol Graph Model* 2004, 22:425–439.

37. Zhou R, Berne BJ, Germain R. The free energy landscape for beta hairpin folding in explicit water. *Proc Natl Acad Sci U S A* 2001, 98:14931–14936.

38. Munoz V. Thompson PA, Hofrichter J, Eaton WA. Folding dynamics and mechanism of beta-hairpin formation. *Nature* 1997, 390:196–199.

39. Munoz V, Henry ER, Hofrichter J, Eaton WA. A statistical mechanical model for beta-hairpin kinetics. *Proc Natl Acad Sci U S A* 1998, 95:5872–5879.

40. Dinner AR, Lazaridis T, Karplus M. Understanding beta-hairpin formation. *Proc Natl Acad Sci U S A* 1999, 96:9068–9073.

41. 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.

42. Jang H, Hall CK, Zhou Y. Thermodynamics and stability of a beta-sheet complex: molecular dynamics simulations on simplified off-lattice protein models. *Protein Sci* 2004, 13:40–53.

43. 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.

44. Ding F, Dokholyan NV. Dynamical roles of metal ions and the disulfide bond in Cu, Zn superoxide dismutase folding and aggregation. *Proc Natl Acad Sci U S A* 2008, 105:19696–19701.

45. Dokholyan NV. What is the protein design alphabet? *Proteins* 2004, 54:622–628.

46. Ding F, Dokholyan NV. Emergence of protein fold families through rational design. *PLoS Comput Biol* 2006, 2:e85.

47. Eddy SR. Non-coding RNA genes and the modern RNA world. *Nat Rev Genet* 2001, 2:919–929.

48. Huttenhofer A, Schattner P. The principles of guiding by RNA: chimeric RNA-protein enzymes. *Nat Rev Genet* 2006, 7:475–482.

49. Doherty EA, Doudna JA. Ribozyme structures and mechanisms. *Annu Rev Biophys Biomol Struct* 2001, 30:457–475.

50. Merino EJ, Wilkinson KA, Coughlan JL, Weeks KM. RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J Am Chem Soc* 2005, 127:4223–4231.

51. Wilkinson KA, Merino, EJ, Weeks KM. Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat Protoc* 2006, 1:1610–1616.

52. Buldyrev SV, Kumar P, Debenedetti PG, Rossky PJ, Stanley HE. Water-like solvation thermodynamics in a spherically symmetric solvent model with two characteristic lengths. *Proc Natl Acad Sci U S A* 2007, 104:20177–20182.

53. Schultz AJ, Hall CK, Genzer J. Computer simulation of block copolymer/nanoparticle composites. *Macromolecules* 2005, 38:3007–3016.

54. Huh J, Ginzburg VV, Balazs AC. Thermodynamic behavior of particle/diblock copolymer mixtures: simulation and theory. *Macromolecules* 2000, 33:8085–8096.

55. Strickland LA, Hall CK, Genzer J. Simulation of mechanically assembled monolayers and polymers in good solvent using discontinuous molecular dynamics. *Macromolecules* 2008, 41:6573–6581.

56. Ding F, Dokholyan NV. Simple but predictive protein models. *Trends Biotechnol* 2005, 23:450–455.