Published in final edited form as:

Phys Chem Chem Phys. 2011 October 14; 13(38): 16890–16901. doi:10.1039/c1cp20752k.

Coarse-grained force field; general folding theory

Adam Liwo¹, Yi He², and Harold A. Scheraga^{2,*}

¹Faculty of Chemistry, University of Gdansk, ul. Sobieskiego 18, 80-952 Gdansk, Poland ²Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, NY 14853-1301, U.S.A

Abstract

We review the coarse-grained UNited RESidue (UNRES) force field for the simulations of protein structure and dynamics, which is being developed in our laboratory over the last several years. UNRES is a physics-based force field, the prototype of which is defined as a potential of mean force of polypeptide chains in water, where all the degrees of freedom except the coordinates of α -carbon atoms and side-chain centers have been integrated out. We describe the initial implementation of UNRES to protein-structure prediction formulated as a search for the global minimum of the potential-energy function and its subsequent molecular dynamics and extensions of molecular-dynamics implementation, which enabled us to study protein-folding pathways and thermodynamics, as well as to reformulate the protein-structure prediction problem as a search for the conformational ensemble with the lowest free energy at temperatures below the folding-transition temperature. Applications of UNRES to study biological problems are also described.

Introduction

Our approach to the protein-folding problem had its origins in both theoretical and experimental work. Early theoretical work focused on the effect of side chain-side chain hydrogen bonds on pK_a 's of ionizable groups,¹ on the mechanism of hydrophobic interactions,² and the interaction between hydrogen bonds and hydrophobic interactions;³ see Figure 1 of reference 3. Experimental observations showed that the ultraviolet absorption spectrum of tyrosine residues of bovine pancreatic ribonuclease A (RNase A) varied with pH, near pH 10 due to ionization of tyrosines,⁴ and near pH 2 due to ionization of carboxyl groups near tyrosine.⁵ A series of biophysical experiments on RNase A, based on these observations, led to the identification of three Tyr...Asp interactions,⁶ later verified by the X-ray structure of RNase A.⁷

Based on this theoretical and experimental work, and with knowledge of the amino acid sequence of RNase A⁸ including the location of its four disulfide bonds, these distance constraints provided the motivation to develop a procedure to compute the three-dimensional structure and folding pathways of a protein by combining these distance constraints with an empirical potential energy function. The methodology evolved from use of a hard-sphere potential function^{9,10} to an all-atom potential, ECEPP, (Empirical Conformational Energy Program for Peptides),¹¹ and subsequently to a coarse-grained UNited-RESidue potential, UNRES,¹² with global optimization (first of potential energy and subsequently of free energy) based on Anfinsen's thermodynamic hypothesis¹³ that a native protein and its surroundings exist in a state of lowest free energy. The theoretical basis for this approach was formulated in references 14, 15, 16 and 17. This article traces the

*Corresponding author; phone: (607) 255 4034, fax: (607) 254 4700, has5@cornell.edu.

evolution of the force field up to UNRES, and discusses recent improvements and applications of UNRES.

The two aspects of a folding algorithm are the potential energy and the procedure for global optimization of the potential energy. Both aspects are closely related. Global optimization has involved energy minimization, Monte Carlo, and Molecular Dynamics, and variants thereof. These have been used as components of two different approaches, a physics-based one in which the structure and folding pathways are determined solely by the potential energy, and a knowledge-based one that makes use of sequence alignments, secondary structure prediction, homology modeling, threading, or fragment coupling. We have used the former, physics-based approach to determine structure and folding pathways in order to gain an understanding as to how physics determines the behavioral aspects of protein structure, folding, and reactivity.

Initially, we used various energy minimization and Monte Carlo procedures to treat small systems involving linear peptides, cyclic peptides (with constraints for exact ring closure), and fibrous proteins. Examples include the cyclic decapeptide gramicidin S,¹⁸ poly-Gly-Pro-Pro¹⁹ a model of collagen, and the globular 46-residue three-helical B domain of staphylococcal protein A.²⁰ More recently, our efforts have been focused on development and application of the coarse-grained UNRES model.²¹

Development of UNRES

UNRES was developed to surmount the limitations in available computer time when using an all-atom potential-energy function to compute protein structure. It was based on averaging out those degrees of freedom that were considered to be non-essential to the attainment of the thermodynamically most-stable structure of a globular protein.^{12,17} In a hierarchical approach to protein-structure prediction, it was originally intended to use CSA,²² a Conformational-Space Annealing method to provide the *region* in which the protein of lowest potential energy would lie. Then, the ensemble of proteins in this smaller *region* would be converted to an all-atom representation which would be searchable with an all-atom potential function. In some applications, the final search with the all-atom potential could be avoided.

The coarse-graining scheme that underpins UNRES defines the effective energy function as a restricted free energy (RFE) function or potential of mean force (PMF) of polypeptide chains in water, in which all degrees of freedom except those defining the shape of a protein have been integrated out.^{12,17} In the UNRES model^{12,17,23-27} two types of interaction sites are defined: side chains (SC) and peptide group (p). This choice was motivated by the fact that the major non-local forces that determine the structure of polypeptide chains involve hydrophobic interactions between side chains, which define the general shape of a protein and hydrogen-bonding interactions that fix the fine details of the structure.²⁸ Thus, a polypeptide chain in UNRES is represented by a sequence of α -carbon (C^α) atoms with united side chains attached to the C^α 's and united peptide groups located halfway between the consecutive C^α atoms (Figure 1). The C^α 's serve only to define the geometry but are not interaction sites; owing to their presence the directionality of peptide-group interactions is accounted for. In the extent of reduction of representation and definition of interaction sites UNRES has much in common with other coarse-grained models of proteins,^{29,30} e.g., the early model of Levitt,³¹ the CABS model developed by Koliński³² (which is based on a coarse-grained protein model developed earlier by Koliński and Skolnick³³) or the MARTINI model.³⁴

The prototype of the UNRES effective energy function (the potential of mean force of polypeptide chains in water) is expressed by eq 1.^{12,17}

$$F(\mathbf{X}) = -RT \left\{ \ln \frac{1}{V_y} \int \exp \left[-\frac{E(\mathbf{X}, \mathbf{y})}{RT} \right] dV_y \right\} \quad (1)$$

where $\mathbf{X}$ is a shorthand for the degrees of freedom that are kept in the coarse-grained model (in UNRES the C^α and SC coordinates), $\mathbf{y}$ denotes those which are averaged out in coarse-graining (the solvent degrees of freedom, the rotation angles of peptide-group planes about the $C^\alpha \dots C^\alpha$ axes, and the internal degrees of freedom of the side chains), E is the potential energy of the system, R is the universal gas constant, and T is the absolute temperature.

An important feature of the effective energy function defined by eq 1 is that the thermodynamic functions of the system under study are the same as those calculated with the all-atom potential-energy function [$E(\mathbf{X}, \mathbf{y})$ in eq 1]. Moreover, $\exp[-\beta F(\mathbf{X})]$ is proportional to the probability density of a coarse-grained conformation to occur.^{12,17,35-38} This principle was applied in a rigorous manner in the force-matching method developed by Voth and coworkers.³⁵⁻³⁷

The PMF defined by eq 1 can be expressed as a sum of cluster cumulant functions or factors,¹⁷ each of which corresponds to interactions within or between a given number of interacting sites. The order of a factor is the number of site-site or inter-site interactions involved. Factors of order 1 consist of single site-site interactions, while factors of higher order are correlation terms. We found¹⁷ that the most important correlation terms are those that couple local conformational states of the backbone with backbone-electrostatic (hydrogen-bonding) interactions. The UNRES energy function is given by eq 2.

$$\begin{aligned} U = & w_{sc} \sum_{i < j} U_{sc,sc_j} + w_{scp} \sum_{i \neq j} U_{sc,p_j} + w_{pp} f_2(T) \sum_{i < j-1} U_{p,p_j} + w_{tor} f_2(T) \sum_i U_{tor}(\gamma_i) \\ & + w_{tord} f_3(T) \sum_i U_{tord}(\gamma_i, \gamma_{i+1}) + w_b \sum_i U_b(\theta_i) + w_{rot} \sum_i U_{rot}(\alpha_{sc_i}, \beta_{sc_i}, \theta_i) \\ & + w_{bond} \sum_i U_{bond}(d_i) + \sum_{m=3}^6 w_{corr}^{(m)} f_m(T) U_{corr}^{(m)} + w_{ss} \sum_i U_{ss,i} \end{aligned} \quad (2)$$

with^{39,40}

$$f_m(T) = \frac{\ln(e + e^{-1})}{\ln \left\{ \exp \left[\left(\frac{T}{T_0} \right)^{m-1} \right] + \exp \left[-\left(\frac{T}{T_0} \right)^{m-1} \right] \right\}}; T_0 = 300K \quad (3)$$

where the successive terms represent side chain-side chain, side chain-peptide, peptide-peptide, torsional, double-torsional, bond-angle bending, side-chain local (dependent on the angles θ , α , and β of Figure 1), distortion of virtual bonds, multi-body (correlation) interactions, and formation of disulfide bonds, respectively. The temperature-dependent factors defined by eq 3^{39,40} reflect the fact that the effective UNRES energy is an approximate cumulant expansion of the restricted free energy (which depends on temperature). In the generalized-cumulant expansion of the RFE, the cumulant of order n is multiplied by $1/T^{n-1}$ as is the asymptotical behavior of eq 3 at high temperatures. Because $1/T^{n-1}$ rapidly grows as the temperature gets smaller, we implement eq 3 to scale the cumulant terms instead of $1/T^{n-1}$.³⁹ The temperature dependence was not introduced in earlier applications which involved a global-energy-minimum search and, thus, assumed that the temperatures was always the physiological temperature. In these applications all of the factors were equal to 1.

Initially,^{24,26} the correlation terms were not present in UNRES and most of the parameters of effective energy expression, except those of U_{ppf} , were determined from the distribution and correlation functions from the Protein Data Bank (PDB). The energy-term weights were determined by optimizing the Z-score function by using a set of energy-minimized decoys (threading with energy minimization)²⁶. Later,^{12,17} we introduced the correlation terms and reparameterized most of the terms based on quantum-mechanical ab initio or semiempirical energy surfaces of model systems.⁴¹⁻⁴⁵

The w 's in eq 1 are the weights of the respective energy terms. They are determined to reproduce the native structures and thermodynamics of folding of selected proteins; this procedure is termed force-field calibration. Initially,^{26,46,47} we maximized the energy gap between the native-like structure and the lowest-energy alternative structure. The procedure consisted of iterations in which a conformational search was performed by using the CSA method and then the gap was evaluated and optimized using the set of decoy structures generated by CSA runs with the current parameters. Iteration terminated when the native-like structures were the lowest-energy ones. This approach worked for small calibration proteins with simple folds but not for more complicated β and $\alpha+\beta$ folds, which reduced the predictive power of the force field. Therefore, later,⁴⁸⁻⁵⁰ we developed a hierarchical-optimization procedure in which the conformational space is divided into levels of conformations with some native-like elements; this degree of native-likeness increases with level number, while the free energy of the level decreases with increasing level number. With this improvement, we produced the 4P force field⁵⁰ parameterized with the following four proteins: 1GAB (α), 1E0L (β), 1E0G ($\alpha+\beta$), and 1IGD ($\alpha+\beta$).

UNRES was also extended to treat oligomeric proteins.⁵¹⁻⁵³ The development of UNRES over the years is summarized in Table 1. A more detailed comparison of UNRES with related coarse-grained approaches is provided in ref. 29, and the theory and development of UNRES is also described in ref 27.

Some Original Applications of UNRES

Initially, UNRES was implemented to find the native structure of a protein as the global minimum of the energy function. First applications of UNRES and the powerful Conformational Space Annealing (CSA) method of global optimization to the N-terminal part of the B-domain of staphylococcal protein A (46 residues, a three- α -helix bundle) and to calbindin (75 residues, a four- α -helix bundle)⁵⁵ encouraged us to participate in the Third Community Wide Experiment on the Critical Assessment of Techniques for Protein Structure Prediction (CASP3), in which we obtained the best prediction of the structure of HDEA (target T0061) among all the groups participating. The quality of the structure was 4.2 Å in C $^{\alpha}$ -distance root-mean-square deviation from the experimental structure and our prediction was wrong only in that the N-terminal α -helix was flipped with respect to that of the experimental structure.^{21,56,57} The predicted structure is superposed on the experimental structure in Figure 2a and 2b.

As the UNRES force field evolved, it was gradually becoming able to predict larger proteins with more complex folds. In CASP4, we predicted, for the first time, the complete structure of a 70-residue protein, bacteriocin S (target T0102); with the introduction of correlation terms, we also were able to treat proteins with β -structure.⁵⁸ During the CASP5 exercise in 2004, we predicted nearly the complete structure (203 out of 235-residues) of a six- α -helix bundle protein (target T0198), complete or nearly complete structures of three other smaller α - and $\alpha+\beta$ -proteins, and significant portions of structures of other proteins.⁵⁹ In these predictions, we used the 4P force field.⁵⁰ Because of our switching to an investigation of protein dynamics and thermodynamics, following the development of Langevin dynamics

and generalized-ensemble techniques for UNRES, we stopped the development of global-optimization-oriented prediction after CASP6.

Simulation of Protein Folding Pathways by Molecular Dynamics with UNRES

With results described in the preceding section, the application of UNRES was extended from computation of protein structure to the determination of folding pathways by molecular dynamics (MD) with UNRES and Lagrange equations of motion for Langevin dynamics in the canonical ensemble, including friction and random forces. It was expected that the limitations of MD with an all-atom force field would be overcome because the fast degrees of motion are averaged out in UNRES.⁶⁰⁻⁶³ Consequently, molecular dynamics with UNRES and its extension (UNRES/MD) were developed.^{60,61} Because the geometry of an UNRES polypeptide chain is not uniquely defined by the Cartesian coordinates of the interacting sites (i.e., the SC and p centers), the Lagrange formalism has been applied to derive the equations of motion for MD calculations with the UNRES model of polypeptide chains using virtual-bond $C^\alpha \dots C^\alpha$ and $C^\alpha \dots SC$ vectors as a set of generalized coordinates. The virtual bonds are represented as elastic rods with uniformly distributed masses.⁶⁰ Non-conservative forces (the friction and random forces) coming from the solvent were included in the framework of Langevin dynamics⁶¹; we also applied the Berendsen thermostat⁶⁴ and later⁶⁵ the Nose-Hoover⁶⁶ and Nose-Poincaré⁶⁷ thermostats to carry out isothermal runs without explicit friction and random forces. A robust algorithm for numerical integration of the equations of motion, termed adaptive multiple-step time-split algorithm (A-MTS) with a velocity-Verlet⁶⁸ integration scheme, was developed⁶⁹ to treat rapid variation of forces during the progress of dynamics, an event which happens frequently in UNRES/MD where large structural changes occur.

The Langevin equation for the UNRES force field is given by eq 4.

$$\mathbf{G} \frac{d^2 \mathbf{q}}{dt^2} = -\nabla_{\mathbf{q}} U(\mathbf{q}) - \mathbf{A}^T \mathbf{\Gamma} \mathbf{A} \frac{d\mathbf{q}}{dt} + \mathbf{A}^T \mathbf{f}^{\text{rand}} \quad (4)$$

where $\mathbf{q} = [\mathbf{dC}_1, \mathbf{dC}_2, \dots, \mathbf{dC}_{n-1}, \mathbf{dX}_1, \mathbf{dX}_2, \dots, \mathbf{dX}_m]^T$ is the vector of virtual $C^\alpha \dots C^\alpha$ ($\mathbf{dC}$) and $C^\alpha \dots SC$ ($\mathbf{dX}$) vectors (n being the number of residues and m the number of non-glycine residues), $\mathbf{G}$ is the (constant) inertia matrix in the $\mathbf{q}$ space, U is the UNRES effective energy, $\mathbf{A}$ is the (constant) matrix transforming the $\mathbf{q}$ coordinates into the Cartesian coordinates of the peptide groups and side-chain centers, $\mathbf{\Gamma}$ is the friction tensor expressed in the Cartesian coordinates of the interacting sites (peptide-group and side-chain centers, respectively), and $\mathbf{f}^{\text{rand}}$ is the vector of random forces acting on interacting sites. Details of the approach can be found in the references cited.⁶⁰⁻⁶² Since the inertia matrix, $\mathbf{G}$, is independent of conformation, its inverse can be computed once and accelerations evaluated by multiplying the forces by its inverse, which spares us costly solution time for a system of linear equations in every MD step.

Because the secondary degrees of freedom are integrated out, the UNRES time scale is wider than that of all-atom molecular dynamics. By comparing the mean-first-passage times (the time required to encounter the folded structure for the first time, τ_f) of the folding of deca-alanine and protein A simulated with the all-atom AMBER force field and UNRES, we found that UNRES provides about a 4-8 time extension of the time scale.⁶¹ Together with the reduction of computational costs, UNRES offers about a 4000-fold extension of the simulation time scale compared to all-atom simulations with explicit solvent.^{61,62} Some other coarse-grained approaches, e.g., that of Koliński, which is based on the CABS high-resolution lattice model with knowledge-based potentials and Monte Carlo dynamics³² have

been developed so far to be applied to carry out unrestricted simulations of real proteins. By using the CABS approach, Koliński and coworkers performed thermal-unfolding simulations of barnase,⁷⁰ chymotrypsin,⁷⁰ and the B1 domain of protein G,⁷¹ obtaining the folding mechanism and intermediate states consistent with experimental data. Because a Monte Carlo approach was used in those simulations, the time scale cannot be defined directly except that it is large enough to cover the folding and unfolding events. Other coarse-grained-dynamics approaches pertain to general protein-like models which capture the general features of folding⁷² or structure-based models designed to locate the native structure of a given protein as the global minimum of their potential-energy surface.⁷³

Apart from protein A and deca-alanine, we carried out a number of successful folding simulations of α -helical proteins and also the α + β -protein 1E0G.⁶² We found that complete folding of a 75-residue protein could be simulated in 4-hour time with a single Athlon processor⁶². With this advantage, we carried out a simulation of the kinetics of folding of staphylococcal protein A.⁶³ We ran 480 canonical molecular dynamics trajectories and collected the statistics from all of them, which allowed us to determine the kinetic equation. Two folding routes were found: a fast one leading directly to the native three-helix bundle and a slow one running through a non-native intermediate (kinetic trap). Later,⁵³ we extended UNRES/MD to treat oligomeric proteins and carried out successful simulations of these proteins.

The initial folding simulations were run with the use of the UNRES force field parameterized for global optimization; the force field was termed 4P because four proteins were used to calibrate it.⁵⁰ We found⁵¹ that, because of the absence of an entropy contribution, the lowest-energy conformations never appear in simulations carried out at room temperature, which puts into question the use of global optimization of the potential energy function in protein-structure prediction. Ignoring conformational entropy in force-field calibration also led to poor performance of the force field on proteins with β -structure in molecular dynamics simulations. Our further efforts focused on reparameterization of UNRES for molecular dynamics simulations.

Conformational ensembles and folding thermodynamics

With the use of MD as a powerful engine in conformational search with UNRES, it also became possible to investigate the thermodynamics of protein folding and conformational ensembles corresponding to different stages of folding. We implemented⁷⁴ Replica Exchange Molecular Dynamics (REMD),⁷⁵ a Multicanonical (MUCA) method also known as entropic sampling,⁷⁶ Replica Exchange Multicanonical (REMUCA), and Multicanonical Replica Exchange (MUCAREM) molecular dynamics⁷⁷ and, later⁷⁸ Multiplexed Replica Exchange Molecular Dynamics (MREMD)⁷⁹ with UNRES. The MREMD method turned out to be the most efficient to use with UNRES. In recognition of the fact that UNRES is a free-energy function, we also introduced temperature dependence of the free-energy terms (eqs 2 and 3); recently,⁸⁰ we also introduced temperature dependence to the $U_{SC_iSC_j}$ potentials to account for temperature dependence of hydrophobic association. The use of MREMD together with the weighted histogram analysis method (WHAM)⁸¹ enabled us to compute thermodynamic properties of proteins at all temperatures.³⁹

With the MREMD method, we were able to calculate the heat-capacity curves. In accordance with our earlier canonical simulations, we found that the 4P force field (designed for global optimization) gave folding-transition temperatures of about 800 K, which is unphysical. This result was expected because, when optimizing the 4P force field, we maximized the energy gaps between the native and alternative structures.

To address the shortcomings of the 4P force field that arose in UNRES/MD simulations, we modified the hierarchical optimization method to include decoys generated with MREMD simulations.^{40,82} Proper folding-transition temperatures were achieved by setting the requirement that the free-energy curves of levels containing native-like or partially native-like conformations intersect with those containing non-native conformations at the folding-transition temperature. We also included experimental free-energy curves in optimization (Figure 3). Finally⁸³, the optimization procedure was improved to search many sets of energy-term weights simultaneously.

With the capacity of UNRES/MREMD to produce ensembles of conformations, we designed a fully physics-based protein structure prediction algorithm, in which the results of a replica-exchange molecular dynamics run are processed by WHAM⁸¹ to calculate the probability of each conformation, then the conformations are ranked according to decreasing probabilities and, finally, those whose probabilities sum up to a cut-off value (usually 0.99) are clustered, and the probability of each cluster is calculated from eq 5.

$$P_{\{l\}} = \frac{\sum_{i \in \{l\}} \exp\left(\omega_i - \frac{U_i}{RT}\right)}{\sum_{\{l\}} \sum_{i \in \{l\}} \exp\left(\omega_i - \frac{U_i}{RT}\right)} \quad (5)$$

where $\{l\}$ denotes the set of conformations belonging to the l th cluster, $P_{\{l\}}$ is the probability of the l th cluster, U_i is the energy of the i th conformation, ω_i is the logarithm of the weight of this conformation determined by the WHAM method,³⁹ R is the universal gas constant, and T is the absolute temperature.

With the new approach, in the CASP9 exercise held in year 2010, for three targets [T0534-D2 (domain 2), T0537-D2, and T0578-D1], the Global Distance Test (GDT) plots of our predictions were ahead of those of other groups (Figure 4). (GDT is the percentage of residues of the largest set, not necessarily continuous, deviating from the experimental structure by no more than a specified distance cutoff.) However, none of these predictions was our first-choice model (model 1) for a particular target, and we also achieved lower rmsd resolution for shorter segments. Nevertheless, it is the first time that our fully physics-based methodology (both as far as the energy function and conformational sampling is concerned) succeeded for more than one target in a CASP exercise. For a number of other targets in CASP9, as well as in CASP7 and CASP8, we also achieved good results but, because they were obtained for homologous targets, these results were worse than those obtained by homology modeling.

Improved Parallelism and Extension to Large Structures

While UNRES provides tremendous reduction of the cost of energy and force evaluation and extension of the time scale in MD simulations, with respect to treating polypeptide chains at the all-atom level, because of reduction of the number of the degrees of freedom, it still takes several days of computations to search the conformational space of about a 200-residue protein with the CSA method or more than a week to run MREMD until convergence of thermodynamic properties (usually about 20 million MD steps/trajectory) with one processor computing energy and forces of a single conformation. Fine-graining (i.e., splitting energy and force evaluation between processors), therefore, becomes necessary to treat larger proteins. Hence, we parallelized energy and force evaluation in UNRES⁸⁴ achieving very good scalability (Figure 5). Because there is no explicit solvent in UNRES, we partitioned the interactions, but did not partition the interaction sites between processors, as commonly applied in the all-atom treatment with explicit solvent. The evaluation of accelerations from forces (eq 4) has also been parallelized.

With fine-grain parallelization, 20 million MD steps of a 200-residue proteins can be run in a single wall-clock day of computations, and it takes about 4 days of computations to run 20 million MD steps for an 800-residue protein provided that sufficient supercomputer resources are available. We took advantage of the fine-grain parallelism of UNRES in the CASP8 and CASP9 exercises when we were able to treat over 300-residue targets by our physics-based prediction approach, as well as in biological applications of UNRES to simulations of PICK1 binding to PDZ⁸⁵ and of the initial phase of Hsp70 chaperone catalytic cycle.⁸⁶

Further Improvements of UNRES

Our work on the development of the UNRES force field is now focused on the replacement of the present $U_{SC;SC}$ terms describing the effective side chain – side chain interaction potentials. The present potentials²⁵ have simple Gay-Berne⁸⁷ functional forms which implies axial symmetry and, consequently, ignores the fact that polar or charged side chains have non-polar necks and polar or charged headgroups. Moreover, the parameters of the present $U_{SC;SC}$ potentials were derived²⁵ from the side-chain side-chain correlation functions computed from protein structures deposited in the Protein Data Bank.

Recently,⁸⁸⁻⁹³ we started to develop new $U_{SC;SC}$ potentials for UNRES, which take into account the amphipolar character of charged and polar side chains. In this model, non-polar side chains are modeled by single interaction sites with axial symmetry, while polar side chain are represented by a non-polar axially-symmetric site and a polar site. The effective energy of interaction between the non-polar sites consists of a Gay-Berne term representing van der Waals interactions and a solvation term representing hydrophobic hydration. To derive the solvation term, we developed⁸⁸ a model in which the density of the solute is represented as a Gaussian and that of water in the first hydration sphere is represented as a Gaussian differential. The effective energy of solvent-mediated interactions was derived by calculating the overlap integrals between the solvent and solute density and that between the solvent density in the solvation spheres, each belonging to a different solute particle. The charged and polar headgroups are represented by a point charge or a point dipole, respectively, and a van der Waals sphere. The effective interaction energy between the head groups involves a Coulombic term (averaged over dipole-moment rotation about the C^α-SC axis if dipoles are involved), a Generalized-Born⁹⁴ term to account for solvent polarization, and a Gaussian-overlap term to account for non-electrostatic contributions to the potential of mean force of solvent-mediated interactions. A polarization term has been introduced to account for the electrostatic component of the interaction energy of the headgroups with non-polar necks.^{89,92,93}

To parameterize the above-mentioned energy expressions, we carried out molecular dynamics simulations with TIP3P water and AMBER force field for models (side-chain atom plus CH₃ group in place of the C^α atom) for pairs of all 20 natural amino acids (210 pairs) and, subsequently, determined the respective potentials of mean force as functions of the distance between centers and principal-axis orientation (three orientation angles). Finally, the parameters of the new $U_{SC;SC}$ energy expressions were obtained by least-squares fitting to the potentials of mean force.^{90,91} As an example, the potentials of mean force corresponding to different orientations of two valine side chains and the fitted PMF are shown in Figure 6.

Biological Applications

Recently^{85,86,95} we used UNRES/MD to investigate the mechanism of fiber assembly, PDZ binding to BAR domain, and the opening cycle of Hsp70 chaperone. These three biological applications of UNRES are summarized briefly in this section.

To study the mechanism of fiber assembly, we investigated the binding of a new A β ₁₋₄₀ peptide unit to stacks composed of 6- and 7-A β ₁₋₄₀-chain templates taken from the NMR study of Tycko et al.⁹⁶ We found that all binding pathways follow a dock-and-lock step, consistent with the experimental results of fibril elongation. In the first step of the assembly, the monomer forms hydrogen bonds with the fibril template along one of the strands in a two-stranded β -hairpin; in the second step, hydrogen bonds are formed between the new strand and the fibril template. A sample trajectory is shown in Figure 7.

The next biological application of UNRES concerned PICK1, which is a key regulator of AMPA (α -amino-3-hydroxy-5-methylisoxazole-4-propionic acid) receptor traffic. A key step of this regulation involves the interactions between the component PDZ and BAR domains. We used⁸⁵ UNRES/MD and UNRES/MREMD to estimate possible binding sites for the PDZ domain of PICK1, PICK1-PDZ, to the homology-modeled crescent-shaped dimer of the PICK1-BAR domain. The MREMD results show that the preferred binding site for the single PDZ domain is the concave cavity of the BAR dimer. A second possible binding site is near the N-terminus of the BAR domain that is linked directly to the PDZ domain (Figure 8).

Subsequent short MD simulations, used to determine how the PICK1-PDZ domain moves to the preferred binding site on the BAR domain of PICK1, revealed that initial hydrophobic interactions drive the progress of the simulated binding. Thus, the concave face of the BAR dimer accommodates the PDZ domain first by weak hydrophobic interactions, and then the PDZ domain slides to the center of the concave face, where more favorable hydrophobic interactions take over.

We also used⁸⁶ UNRES/MD to study the dynamics of a conformational transition of an Hsp70 (heat-shock protein 70) chaperone. Hsp70s are key molecular chaperones which assist in the folding and refolding/disaggregation of proteins. They consist of a nucleotide-binding domain (NBD) and a substrate-binding domain (SBD), occur in two major conformations having (a) a closed SBD, in which the two domains do not interact, (b) an open SBD in which the two domains interact strongly. In the SBD-closed conformation, SBD is bound to a substrate protein with release occurring after transition to the open conformation. By using UNRES/MD and a nucleotide-free closed form of *E. coli* DnaK as a model, we successfully simulated the spontaneous transition from the SBD-closed to the SBD-open conformation in which SBD interacts strongly with the NBD (Figure 9). The key event was found to be the rotation of the two halves of the NBD towards each other, which initiates a cascade of conformational changes that make the flexible linker, connecting the NBD with the SBD, bind to the NBD, thus initiating interdomain communication. To the best of our knowledge, this is the first successful simulation of the full transition from the SBD-closed to the SBD-open conformation of an Hsp70.

Free-energy landscape and folding mechanism

With the advantage of long time-scale simulations, we carried out detailed investigations of the free-energy landscapes (FELs) of two small proteins: the Fbp 28 WW domain (PDB ID: 1E0L)⁹⁷⁻¹⁰¹ and the N-terminal part of the B-domain of staphylococcal protein A (PDB ID: 1BDD)⁹⁹ by means of principal-component analysis (PCA). We found that three dimensions are required for sufficient description;^{98,99} consequently, using two order parameters (e.g., the rmsd from the experimental structure and the radius of gyration, as commonly applied) might lead to a false description of the features of the free-energy landscapes of proteins. Similarly, monitoring the variation of rmsd from the experimental structure might lead to false impressions that some of the folding trajectories are very similar, while detailed PCA reveals that they, in fact, correspond to completely different folding routes.⁹⁹ Finally, we

concluded that PCA analysis in backbone-internal coordinates reveals more fine details of the FELs than that in backbone-Cartesian coordinates.⁹⁹

Just recently,¹⁰⁰ by using MREMD with UNRES, we performed a detailed analysis of equilibrium ensembles of protein A at various temperatures. We found that, at the folding-transition temperature, the equilibrium ensemble is not a mixture of about 50% folded and 50% unfolded structures but a collection of essentially unfolded structure with side-chain - side-chain contacts close to native contacts. (Figure 10). This result suggests that the folding transition occurs in each protein molecule rather than in a manner that 50% of the ensemble contains completely folded and 50% completely unfolded protein molecules. Despite its fuzziness, the folding transition is manifested by a sharp peak in the heat capacity and results in bimodal distribution of the distances between helices at the folding-transition temperature. We also found that the long-range side-chain – side chain contacts still hold at temperatures that are 20 K higher than the folding-transition temperature.

Conclusions

The coarse-grained UNRES force field developed in our laboratory was initially intended as a crude tool with which to scan the conformational space of polypeptide chains in search of the global energy minimum. With its subsequent definition as a potential of mean force of a polypeptide in water, factorization of the PMF into parts coming from smaller parts of the system and identification of the factors with specific terms in the effective energy function,¹⁷ it has become a physics-based coarse-grained force field with clearly defined and understood terms. Implementation of MD^{60,61} and its extensions⁷⁴⁻⁷⁸ has enabled us to study folding trajectories and thermodynamics of protein folding, as well as develop a fully physics-based protein-structure prediction method, the performance of which is improving in consecutive CASP exercises, although it will probably take a while for it to compete with bioninformatics-based approaches to protein-structure prediction. However, after extension of the size scale achieved by parallelizing the computation of energy and forces, UNRES/MD has already become a powerful tool with which to study protein-folding dynamics and thermodynamics with application to biological processes.

Owing to the enhanced timescale of simulations provided by UNRES, we were able to examine in detail the folding of small single-domain proteins at the ensemble level, which enabled us to simulate folding kinetics. This approach resulted in further development of the methods of analysis of free-energy landscapes of proteins based on principal-component analysis.⁹⁷⁻⁹⁹ Recently,¹⁰⁰ we also challenged the old view of the protein-folding process by showing that the folding transition seems to occur in every molecule and, consequently, that the protein ensemble is not a 50%-50% mixture of completely folded and completely unfolded structures at the folding-transition temperature.

We continue to develop the UNRES force field; at present we replace the last knowledge-based terms ($U_{SC,SC}$) with physics-based terms determined by MD simulations.⁸⁸⁻⁹³ Other improvements will include the treatment of cis-trans isomerization of peptide groups, better treatment of local interactions, as well as the development of conformational-search techniques (force-biased Monte Carlo, high-directional Monte Carlo, etc.), and advanced techniques for efficient simulations of protein-folding kinetics (such as, e.g., the kinetic network models).

Acknowledgments

This work was supported by grants from the National Institutes of Health (GM-14312), the National Science Foundation (MCB05-41633), and the Polish Ministry of Science and Higher Education (N N204 049035). Computational resources were provided by (a) the Argonne Leadership Computing Facility at Argonne National

Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under contract DE-AC02-06CH11357, (b) the John von Neumann Institute for Computing at the Central Institute for Applied Mathematics, Forschungszentrum Jülich, Germany, (c) the National Science Foundation at the National Institute for Computational Sciences (<http://www.nics.tennessee.edu/>), (d) the Informatics Center of the Metropolitan Academic Network (IC MAN) in Gdansk (e) our 624-processor Beowulf cluster at the Baker Laboratory of Chemistry, Cornell University and our 45-processor Beowulf cluster at the Faculty of Chemistry, University of Gdańsk.

References

1. Laskowski M Jr, Scheraga HA. *J Am Chem Soc.* 1954; 76:6305.
2. Némethy G, Scheraga HA. *J Phys Chem.* 1962; 66:1773. Erratum. *J Phys Chem.* 1963; 67:2888.
3. Némethy G, Steinberg IZ, Scheraga HA. *Biopolymers.* 1963; 1:43.
4. Tanford C, Hauenstein JD, Rands DG. *J Am Chem Soc.* 1955; 77:6409–6413.
5. Scheraga HA. *Biochim Biophys Acta.* 1957; 23:196–197. [PubMed: 13412697]
6. Scheraga HA. *Fed Proc.* 1967; 26:1380–1387. [PubMed: 6051318]
7. Wodawer A, Svenson LA, Sjölin L, Gilliland GL. *Biochemistry.* 1988; 27:2705–2717. [PubMed: 3401445]
8. Hirs CHW, Moore S, Stein WH. *J Biol Chem.* 1960; 235:633–647. [PubMed: 14401957]
9. Némethy G, Scheraga HA. *Biopolymers.* 1965; 3:155–184.
10. Scheraga HA. *Adv Phys Org Chem.* 1968; 6:103–184.
11. Momany FA, McGuire RF, Burgess AW, Scheraga HA. *J Phys Chem.* 1975; 79:2361–2381.
12. Liwo A, Kazmierkiewicz R, Czaplewski C, Groth M, Oldziej S, Wawak RJ, Rackovsky S, Pincus MR, Scheraga HA. *J Comput Chem.* 1998; 19:259–276.
13. Anfinsen CB. *Science.* 1973; 181:223–230. [PubMed: 4124164]
14. Gibson KD, Scheraga HA. *Physiol Chem and Phys.* 1969; 1:109–126.
15. G N, Scheraga HA. *J Chem Phys.* 1969; 51:4751–4767.
16. G N, Scheraga HA. *Macromolecules.* 1976; 9:535–542.
17. Liwo A, Czaplewski C, Pillardy J, Scheraga HA. *J Chem Phys.* 2001; 115:2323–2347.
18. Dygert M, Go N, Scheraga HA. *Macromolecules.* 1975; 8:750–761. [PubMed: 53324]
19. Vila JA, Ripoll DR, Scheraga HA. *Proc Natl Acad Sci U S A.* 2003; 100:14812–14816. [PubMed: 14638943]
20. Miller MH, Scheraga HA. *J Polymer Sci: Polymer Symposia.* 1976; (No 54):171–200.
21. Liwo A, Lee J, Ripoll DR, Pillardy J, Scheraga HA. *Proc Natl Acad Sci USA.* 1999; 96:5482–5485. [PubMed: 10318909]
22. Lee J, Scheraga HA, Rackovsky S. *J Comput Chem.* 1997; 18:1222–1232.
23. Liwo A, Pincus MR, Wawak RJ, Rackovsky S, Scheraga HA. *Protein Sci.* 1993; 2:1715. [PubMed: 8251944]
24. Liwo A, Pincus MR, Wawak RJ, Rackovsky S, Scheraga HA. *Protein Sci.* 1993; 2:1715. [PubMed: 8251944]
25. Liwo A, Oldziej S, Pincus MR, Wawak RJ, Rackovsky S, Scheraga HA. *J Comput Chem.* 1997; 18:849.
26. Liwo A, Pincus MR, Wawak RJ, Rackovsky S, Oldziej S, Scheraga HA. *J Comput Chem.* 1997; 18:874.
27. Liwo, A.; Czaplewski, C.; Oldziej, S.; Rojas, AV.; Kazmierkiewicz, R.; Makowski, M.; Murarka, RK.; Scheraga, HA. Coarse-graining of condensed phase and biomolecular systems. Voth, G., editor. CRC Press, Taylor & Francis; Farmington, CT: 2008. p. 107–122.
28. Dill KD. *Biochemistry.* 1990; 29:7133. [PubMed: 2207096]
29. Czaplewski C, Liwo A, Makowski M, Oldziej S, Scheraga HA. Multiscale approaches to protein modeling. :35–83.
30. Tozzini V. *Quart Rev Biophys.* 2010; 43:333.
31. Levitt M. *J Mol Biol.* 1976; 104:59. [PubMed: 957439]
32. Koliński A. *Acta Biochim Pol.* 2004; 51:349. [PubMed: 15218533]

33. Koliński A, Skolnick J. Proteins: Struct Funct Genet. 1994; 18:338. [PubMed: 8208726]
34. Monticelli L, Kandasamy SK, Periole X, Larson RG, Tieleman DP, Marrink SJ. J Chem Theor Comput. 2008; 4:819.
35. Izvekov S, Voth GA. J Phys Chem B. 2005; 109:2469. [PubMed: 16851243]
36. Noid WG, Chu J-W, Ayton GS, Krishna V, Izvekov S, Voth GA, Das A, Andersen HC. J Chem Phys. 2008; 128:244144.
37. Zhou J, Thorpe I, Izvekov S, Voth GA. Biophys J. 2007; 92:4289. [PubMed: 17400700]
38. Voth, GA., editor. Coarse graining of condensed phase and biomolecular systems. CRC Press, Taylor & Francis; Farmington, CT: 2008.
39. Liwo A, Khalili M, Czaplewski C, Kalinowski S, Ołdziej S, Wachucik K, Scheraga HA. J Phys Chem B. 2007; 111:260–285. [PubMed: 17201450]
40. Shen H, Liwo A, Scheraga HA. J Phys Chem B. 2009; 113:8738. [PubMed: 19480420]
41. Ołdziej S, Kozłowska U, Liwo A, Scheraga HA. J Phys Chem A. 2003; 107:8035.
42. Liwo A, Ołdziej S, Czaplewski C, Kozłowska U, Scheraga HA. J Phys Chem B. 2004; 108:9421.
43. Kozłowska U, Liwo A, Scheraga HA. J Phys: Cond Matter. 2007; 19:285203.
44. Kozłowska U, Liwo A, Scheraga HA. J Comput Chem. 2010; 31:1143. [PubMed: 20073062]
45. Kozłowska U, Maisuradze GG, Liwo A, Scheraga HA. J Comput Chem. 2010; 31:1154. [PubMed: 20017135]
46. Lee J, Ripoll DR, Czaplewski C, Pillardy J, Wedemeyer WJ, Scheraga HA. J Phys Chem B. 2001; 105:7291.
47. Pillardy J, Czaplewski C, Liwo A, Wedemeyer WJ, Lee J, Ripoll DR, Arłukowicz P, Ołdziej S, Arnautova Y, Scheraga HA. J Phys Chem B. 2001; 105:7299.
48. Liwo A, Arłukowicz P, Czaplewski C, Ołdziej S, Pillardy J, Scheraga HA. Proc Natl Acad Sci U S A. 2002; 99:1937. [PubMed: 11854494]
49. Ołdziej S, Liwo A, Czaplewski C, Pillardy J, Scheraga HA. J Phys Chem B. 2004; 108:16394.
50. Ołdziej S, Łagiewka J, Liwo A, Czaplewski C, Chinchio M, Nanas M, Scheraga HA. J Phys Chem B. 2004; 108:16950.
51. Saunders JA, Scheraga HA. Biopolymers. 2003; 68:300. [PubMed: 12601791]
52. Saunders JA, Scheraga HA. Biopolymers. 2003; 68:318. [PubMed: 12601792]
53. Rojas AV, Liwo A, Scheraga HA. J Phys Chem B. 2007; 111:293–309. [PubMed: 17201452]
54. Li Z, Scheraga HA. Proc Natl Acad Sci USA. 1987; 84:6611. [PubMed: 3477791]
55. Lee J, Liwo A, Scheraga HA. Proc Natl Acad Sci USA. 1999; 96:2025. [PubMed: 10051588]
56. Lee J, Liwo A, Ripoll DR, Pillardy J, Scheraga HA. Proteins: Struct Funct Genet. 1999; Suppl 3:204.
57. Lee J, Liwo A, Ripoll DR, Pillardy J, Saunders JA, Gibson KD, Scheraga HA. Int J Quant Chem. 2000; 77:90.
58. Pillardy J, Czaplewski C, Liwo A, Lee J, Ripoll DR, Kamińskiewicz R, Ołdziej S, Wedemeyer WJ, Gibson KD, Arnautova YA, Saunders J, Ye Y-J, Scheraga HA. Proc Natl Acad Sci USA. 2001; 98:2329. [PubMed: 11226239]
59. Ołdziej S, Czaplewski C, Liwo A, Chinchio M, Nanas M, Vila J, Khalili M, Arnautova YA, Jagielska A, Makowski M, Schafroth HD, Kamińskiewicz R, Ripoll DR, Pillardy J, Saunders J, Kang Y, Gibson K, Scheraga HA. Proc Natl Acad Sci U S A. 2005; 102:7547. [PubMed: 15894609]
60. Khalili M, Liwo A, Rakowski F, Grochowski P, Scheraga HA. J Phys Chem B. 2005; 109:13785–13797. [PubMed: 16852727]
61. Khalili M, Liwo A, Jagielska A, Scheraga HA. J Phys Chem B. 2005; 109:13798–13810. [PubMed: 16852728]
62. Liwo A, Khalili M, Scheraga HA. Proc Natl Acad Sci U S A. 2005; 102:2362. [PubMed: 15677316]
63. Khalili M, Liwo A, Scheraga HA. J Mol Biol. 2006; 355:536–547. [PubMed: 16324712]
64. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. J Chem Phys. 1984; 81:3684.

65. Kleinerman DS, Czaplewski C, Liwo A, Scheraga HA. J Chem Phys. 2008; 128:245103. [PubMed: 18601387]
66. Hoover G. Phys Rev A. 1985; 31:1695. [PubMed: 9895674]
67. Nose S. J Phys Soc Jpn. 2001; 70:75.
68. Swope WC, Andersen HC, Berens PH, Wilson KR. J Chem Phys. 1982; 76:637.
69. Rakowski F, Grochowski P, Lesyng B, Liwo A, Scheraga HA. J Chem Phys. 2006; 125:204107. [PubMed: 17144690]
70. Kmiecik S, Koli ski A. Proc Natl Acad Sci USA. 2007; 104:12330. [PubMed: 17636132]
71. Kmiecik S, Koli ski A. Biophys J. 2008; 94:726. [PubMed: 17890394]
72. Klimov DK, Thirumalai D. J Chem Phys. 1998; 109:4119.
73. Hills RD, Brooks CL. Int J Mol Sci. 2009; 10:889. [PubMed: 19399227]
74. Nancias M, Czaplewski C, Scheraga HA. J Chem Theor Comput. 2006; 2:513.
75. Hansmann UHE. Chem Phys Lett. 1997; 1-3:140.
76. Lee J. Phys Rev Lett. 1993; 71:211. [PubMed: 10054892]
77. Mitsutake A, Sugita Y, Okamoto Y. J Chem Phys. 2003; 118:6664.
78. Czaplewski C, Kalinowski S, Liwo A, Scheraga HA. J Chem Theor Comput. 2009; 5:627.
79. Rhee YM, Pande VS. Biophys J. 2003; 2:775. [PubMed: 12547762]
80. Ołdziej S, Czaplewski C, Liwo A, Scheraga HA. 2010 IEEE International Conference on Bioinformatics and Bioengineering. 2010:263–266.
81. Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JM. J Comput Chem. 1992; 8:1011.
82. Liwo, A.; Czaplewski, C.; Ołdziej, S.; Kozłowska, U.; Makowski, M.; Kalinowski, R.; Ka mierzewicz, R.; Shen, H.; Maisuradze, G.; Scheraga, HA. NIC Series, NIC Symposium 2008. Munster, G.; Wolf, D.; Kremer, M., editors. Vol. 39. 2008. p. 63
83. He Y, Xiao A, Liwo A, Scheraga HA. J Comput Chem. 2009; 30:2127–2135. [PubMed: 19242966]
84. Liwo A, Ołdziej S, Czaplewski C, Kleinerman DS, Blood P, Scheraga HA. J Chem Theory and Comput. 2010; 6:890–909. [PubMed: 20305729]
85. He Y, Liwo A, Weinstein H, Scheraga HA. J Mol Biol. 2011; 405:298–314. [PubMed: 21050858]
86. Goła E, Czaplewski C, Ołdziej S, Maisuradze GG, Senet P, Scheraga HA, Liwo A. J Chem Theory Comput. submitted.
87. Gay JG, Berne BJ. J Chem Phys. 1981; 74:3316.
88. Makowski M, Liwo A, Scheraga HA. J Phys Chem B. 2007; 111:2910. [PubMed: 17388416]
89. Makowski M, Liwo A, Maksimiak K, Makowska J, Scheraga HA. J Phys Chem B. 2007; 111:2917. [PubMed: 17388417]
90. Makowski M, Sobolewski E, Czaplewski C, Liwo A, Ołdziej S, No JH, Scheraga H. J Phys Chem B. 2007; 111:2925. [PubMed: 17388418]
91. Makowski M, Sobolewski E, Czaplewski C, Ołdziej S, Liwo A, Scheraga HA. J Phys Chem B. 2008; 112:11385. [PubMed: 18700740]
92. Makowski M, Liwo A, Sobolewski E, Scheraga HA. J Phys Chem B. 2011 ASAP article. 10.1021/jp111258p
93. Makowski M, Liwo A, Scheraga HA. J Phys Chem B. 2011 ASAP article. 10.1021/jp111259e
94. Still WC, Tempczyk A, Hawley RC, Hendrickson T. J Am Chem Soc. 1990; 112:6127.
95. Rojas A, Liwo A, Browne D, Scheraga HA. J Mol Biol. 2010; 404:537–552. [PubMed: 20888834]
96. Tycko R. Q Rev Biophys. 2006; 39:1. [PubMed: 16772049]
97. Maisuradze GG, Liwo A, Scheraga HA. J Mol Biol. 2009; 385:312. [PubMed: 18952103]
98. Maisuradze GG, Liwo A, Scheraga HA. Phys Rev Lett. 2009; 102:238102. [PubMed: 19658975]
99. Maisuradze GG, Liwo A, Scheraga HA. J Chem Theor Comput. 2010; 6:583.
100. Maisuradze GG, Liwo A, Ołdziej S, Scheraga HA. J Am Chem Soc. 2010; 132:9444. [PubMed: 20568747]
101. Maisuradze GG, Senet P, Czaplewski C, Liwo A, Scheraga HA. J Phys Chem A. 2010; 114:4471. [PubMed: 20166738]

Figure 1.

The UNRES model of polypeptide chains. The interaction sites are peptide-bond centers (p), and side-chain ellipsoids of different sizes (SC) attached to the corresponding α -carbons with different “bond lengths”, b_{SC} . The α -carbon atoms are represented by small open circles. The equilibrium distance of the $C^\alpha \dots C^\alpha$ virtual bonds is taken as 3.8 Å, which corresponds to planar *trans* peptide groups. The geometry of the chain can be described either by the virtual-bond vectors dC_i ($C^\alpha_{i-1} \dots C^\alpha_{i+1}$), $i=1,2,\dots,n-1$ and dX_i ($C^\alpha_i \dots SC_i$), $i=2,3,\dots,n-1$ (represented by thick dashed arrows), where n is the number of residues, or in terms of virtual-bond lengths, backbone virtual-bond angles θ_i , $i=1,2,\dots,n-2$, backbone virtual-bond-dihedral angles γ_i , $i=1,2,\dots,n-3$, and the angles α_i and β_i , $i=2,3,\dots,n-1$ that describe the location of a side chain with respect to the coordinate frame defined by C^α_{i-1} , C^α_i and C^α_{i+1} . Reproduced with permission from Figure 1 of ref 84.

Figure 2.

Model 1 of the CASP3 target T0061 obtained with the UNRES force field and CSA global-optimization procedure (yellow ribbon) superposed on monomer A (red ribbon) of the experimental structure of HDEA. Misaligned helices H-1 and H-2 of model 1 occupy the place of helix H-3 of monomer B (yellow ribbon) of the native dimer. Reproduced with permission from Figures 1 and 2 of ref 21.

Figure 3. (Top) Experimental (filled circles) and calculated with the optimized force field (solid line) free-energy difference between the folded and unfolded state of the 1ENH protein. (bottom) Calculated (solid line) and experimental (dashed line) heat-capacity curve of 1ENH. Data from ref 65; reproduced with permission from Figures 6a and 6b of that reference.

Figure 4. Global Distance Test (GDT) plots of our models of CASP9 targets compared with those of the models from other groups for targets (a) T0534-D2 (b) T0537-D2 and (c) T0578-D1. Red lines represent our models 1, dark-brown lines represent our other models (except models 1), and orange lines represent models from other groups. From the CASP9 web page (<http://predictioncenter.org/casp9/results.cgi>).

Figure 5. Scalability plots of canonical MD simulations of proteins with different sizes obtained with IBM BlueGene (intrepid.alcf.anl.gov). Reproduced with permission From Figure 7c of ref 67.

Figure 6.

The PMF curves for a pair of isobutane molecules (to model a pair of valine side chains). The dashed, dotted, and dot-dashed lines correspond to PMFs determined for the side-to-side, edge-to-edge, and side-to-edge orientation, respectively, obtained by MD simulations. The solid lines correspond to the analytical approximation to the PMFs, with coefficients determined by least-squares fitting (73) of the analytical expression to the PMF determined by MD simulations. Reproduced with permission from Figure 4c of ref 73.

Figure 7. Snapshots of a sample trajectory of A β ₁₋₄₀ fibril propagation. Reproduced with permission from Figure 7 of ref 78.

Figure 8.

Representative structures of PICK1 (blue) bound to the BAR domain obtained in UNRES/MREMD simulations (colored from purple to red from the N- to the C-terminus). Reproduced with permission from Figure 7 of ref 68.

Figure 9.

Snapshots from the canonical MD trajectory of 2KHO run with the UNRES force field. The chain is colored blue to red from the N- to the C-terminus. Reproduced with permission from Figure 2 of 69.

Figure 10.

Experimental structure of 1BDD (a) and four representatives from the canonical ensemble determined by MREMD simulations at T = 325 K (b-e) with nonpolar residues (Ala13, Phe14, Ile17, Leu18, Phe31, Ile32, Leu35, Leu45, Leu46, Ala49, and Leu52) that make contact between helices in the N-terminal and the middle or the N-terminal and the C-terminal α -helix in the experimental structure and in the conformations of the ensemble at T = 325 K, shown as bonds between heavy atoms for the experimental structure or as spheres at the united side-chain centers for the UNRES structures. The residues of the N-terminal, middle, and C-terminal helices are colored blue, green, and red, respectively. Native contacts between the above-mentioned nonpolar residues are shown as black dashed lines, and non-native contacts (in panels b-e) are shown as pink dashed lines. It is shown that a cluster composed of all or some of the above-mentioned nonpolar residues is preserved regardless of the formation of secondary structure or topology which is similar to that of the native structure in conformations (b) and (c) or mirror image (d) and (e). Additionally, in the structures shown in panel (b) and (d), the neighboring nonpolar residues that are not involved in the formation of long-range contacts in the experimental structure: Ala43 and Tyr15, respectively, can assist in the formation of long-range contacts between the N-terminal, middle, and C-terminal segments of the chain. Reproduced with permission from Figure 4 of ref 83.

Table 1

Chronology of most important developments of UNRES.

Time span	Model and force-field features	Conformational search techniques	Applications
1993-1997	Model defined. Part of energy expression defined by Boltzmann averaging of the all-atom potential energy, part as knowledge-based potentials (contact energies) taken from the literature. ²³	Monte Carlo plus Energy Minimization (MCM). ^{23,54}	Avian pancreatic polypeptide – native structure located as global minimum of the potential energy function. ²³
1997-1999	Anisotropic (Gay-Berne) side chain – side chain interaction potentials; knowledge-based parameterization of side-chain and local potentials. ^{25,26} Cooperative terms in backbone-peptide-group interactions introduced. ¹²	Conformational Space Annealing (CSA). ²²	Native structures of many α -helical proteins located as global minima in the UNRES potential energy surface. First successful blind prediction in the CASP3 experiment. ²¹
1999-2002	The force field rigorously defined as a restricted free-energy function (eq 1); factor expansion of the RFE introduced and correlation terms derived based on generalized-cumulant expansion. Calibration by energy-gap maximization. ¹⁷ Extension of the model to multichain proteins. ^{51,52}		Prediction of the structure of proteins of all structural classes by search of the global minimum of the potential energy function. ^{58,59}
2002-2004	Hierarchical optimization of the energy function introduced. ⁴⁸⁻⁵⁰		
2003-2005	Torsional, double-torsional, backbone-electrostatic, and correlation terms parameterized based on quantum-mechanical ab initio calculations. ^{41,42}		
2005-2010	Virtual-bond valence and side-chain-rotamer potentials determined from quantum-mechanical calculations. ⁴³⁻⁴⁵	Coarse-grained molecular dynamics and its extensions. ^{60-62,65,69,74,78} Extension of the MD treatment to multichain proteins. ⁵³	Folding pathways, ^{62,85,86} kinetics, ⁶³ free-energy landscapes, ⁸⁷⁻¹⁰¹ and thermodynamics ^{39,82} of protein folding and conformational changes; physics-based prediction of protein structure (search for the most probable conformational ensembles). ³⁹
2007-2009	Temperature dependence of the effective energy function. ^{39,40} Calibration based on protein-folding thermodynamic data. ^{39,82}		
2005-present	Derivation of physics-based side chain-side chain interaction potentials. ⁸⁸⁻⁹³		