John von Neumann Institute for Computing

Mesoscopic Dynamics with the UNRES Force Field – a Tool for Studying the Kinetics and Thermodynamics of Protein Folding

A. Liwo, C. Czaplewski, S. Ołdziej, M. Chinchio,
A. V. Rojas, M. Khalili, M. Makowski, S. Kalinowski,
U. Kozłowska, R. K. Murarka, H. A. Scheraga

http://www.fz-juelich.de/nic-series/volume36

Mesoscopic Dynamics with the UNRES Force Field – a Tool for Studying the Kinetics and Thermodynamics of Protein Folding

Adam Liwo[1,2], **Cezary Czaplewski**[1,2], **Stanisław Ołdziej**[1,2], **Maurizio Chinchio**[2], **Ana V. Rojas**[2,3,4], **Mey Khalili**[5], **Mariusz Makowski**[1,2], **Sebastian Kalinowski**[1], **Urszula Kozłowska**[2], **Rajesh K. Murarka**[2], **and Harold A. Scheraga**[2]

[1] Faculty of Chemistry, University of Gdańsk, Sobieskiego 18, 80-952 Gdańsk, Poland
E-mail: {adam, czarek, stan, momo, bes}@chem.univ.gda.pl

[2] Baker Laboratory of Chemistry, Cornell University, Ithaca, NY 14843-1301, USA
E-mail: {m.chinchio, uad2, rkm25, has5}@cornell.edu

[3] Department of Physics and Astronomy and

[4] Center for Computation and Technology, Louisiana State University,
Baton Rouge, LA 70803-4001, USA
E-mail: arojas1@lsu.edu

[5] University Chemical Laboratories, Lensfield Road, Cambridge, UK CB2 1EW
E-mail: mk520@cam.ac.uk

All-atom simulations of protein folding starting from arbitrary structures are currently possible only for small proteins and, consequently, united-residue models of polypeptide chains are used in the field. The UNRES model and the respective force field developed in our laboratory belong to this class. As opposed to knowledge-based or heuristic coarse-grained force fields, UNRES was carefully derived as a restricted free energy function of a polypeptide chain, in which fine degrees of freedom not included in the model have been integrated out. This article summarizes recent developments and applications of the force field.

1 Introduction

United-residue or mesoscopic models of proteins receive considerable attention because, in contrast to all-atom simulations, they enable us to study the folding of large proteins at micro- or millisecond time scale. Most of the existing approaches are knowledge-based, which makes explicit use of the information from protein structural databases during the process of conformational search. These approaches are very successful in predicting the structure of proteins[1] but their application to study folding dynamics is limited.

During the last few years we have been developing a mesoscopic physics-based force field termed UNRES (UNited RESidue)[2]. Its initial applications were energy-based predictions of protein structure defined as the lowest-energy conformation[2,3]. Despite considerable success in this field, we realized that such energy-based prediction does not correspond to the process of formation of the native structure, which, following Anfinsen's thermodynamic hypothesis[4] is a collection of very similar conformations occupying the basin with the lowest free energy and, consequently, we changed the approach to ensemble-based global optimization. We also extended the application of UNRES to study the thermodynamics, dynamics, and kinetics of protein folding.

2 The UNRES Force Field

In the UNRES model a polypeptide chain is represented as a sequence of α-carbon atoms (C^α) with attached united side chains (SC) and united peptide groups (p), each of which is positioned in the middle between two consecutive C^α atoms, as shown in Figure 1.
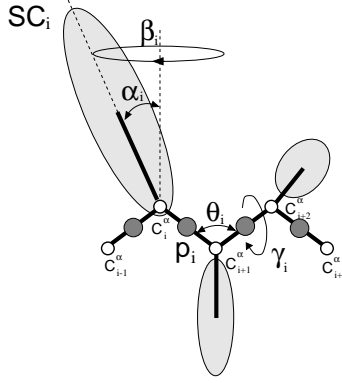

Figure 1. The UNRES model of the polypeptide chain. Dark circles represent united peptide groups (p), open circles represent the C^α atoms, which serve as geometric points. Ellipsoids represent side chains, with their centers of mass at the SC's. The p's are located half-way between two consecutive C^α atoms. The virtual-bond angles θ, the virtual-bond dihedral angles γ, and the angles α_{SC} and β_{SC} that define the location of a side chain with respect to the backbone are also indicated.

The effective energy function is a sum of different terms corresponding to interactions between the SC ($U_{SC_iSC_j}$), SC and p ($U_{SC_ip_j}$), and p ($U_{p_ip_j}$) sites, as well as local terms corresponding to bending of virtual-bond angles θ (U_b), side-chain rotamers (U_{rot}), virtual-bond torsional (U_{tor}) and double-torsional (U_{tord}) terms, virtual-bond-stretching (U_{bond}) terms, correlation terms ($U_{corr}^{(m)}$) pertaining to coupling between backbone-local and backbone-electrostatic interactions[5] (where m denotes the order of correlation), and a term accounting for the energetics of disulfide bonds (U_{SS}). Each of these terms is multiplied by an appropriate weight, w. The energy function is given by equation 1.

$$
\begin{aligned}
U = &\, w_{SC} \sum_{i<j} U_{SC_iSC_j} + w_{SCp} \sum_{i \neq j} U_{SC_ip_j} + w_{pp} \sum_{i<j-1} U_{p_ip_j} \\
&+ w_{tor} \sum_i U_{tor}(\gamma_i) + w_{tord} \sum_i U_{tord}(\gamma_i, \gamma_{i+1}) \\
&+ w_b \sum_i U_b(\theta_i) + w_{rot} \sum_i U_{rot}(\alpha_{SC_i}, \beta_{SC_i}) + \sum_{m=3}^{6} w_{corr}^{(m)} U_{corr}^{(m)} \\
&+ w_{bond} \sum_{i=1}^{nbond} U_{bond}(d_i) + w_{SS} \sum_i U_{SS;i}
\end{aligned} \tag{1}
$$

The prototype of equation 1 is the restricted free energy (RFE) corresponding to a given coarse-grained conformation of a polypeptide chain in water in which all secondary

degrees of freedom, such as the solvent coordinates, internal rotation angles of the side chains, and the angles for the rotation of the peptide groups about the $C^\alpha \cdots C^\alpha$ virtual bonds have been integrated out[5]. We expressed the RFE[5] in terms of Kubo's cluster cumulants[6] and the cluster cumulants into a generalized cumulant series. The fact that the UNRES energy function has the sense of a free energy implies its dependence on temperature, which we introduced recently[7] by multiplying the energy-term weights by appropriate temperature factors.

3 Search of Conformational Space with UNRES

In our earlier implementations of UNRES[2], we were looking for the global minimum of the energy. We developed a global-optimization tool termed a Conformational Space Annealing (CSA) method[8], based on the framework of genetic algorithms. This approach produces the lowest energy minima and can be identified with finding a "folded" structure at the temperature of 0 K. This approach ignores the conformational entropy and, consequently, some proteins whose native-like structures were found as lowest in energy in CSA searches failed to fold to the native structures in canonical MD simulations with the UNRES force field[9].

To obtain a protocol closer to the real folding process, we developed[9] mesoscopic molecular dynamics, in which the UNRES energy function (equation 1) plays the role of the potential energy; we refer to this approach as UNRES/MD. We use either the equations of motion with explicit stochastic and friction terms to account for non-conservative forces coming from the solvent (Langevin dynamics; appropriate to study folding dynamics) or couple the Newton-like equations of motion with the Berendsen thermostat, which results in faster search of the conformational space. Recently[10], we extended the UNRES/MD approach to study the folding of multichain proteins and also developed a procedure to handle dynamic formation and breaking of disulfide bonds during UNRES/MD runs[11]. For better search of the conformational space we implemented[12] the multiplexing replica-exchange method[13] in UNRES/MD (this algorithm is referred to as UNRES/MREMD). We parallelized UNRES/MREMD to run on sytems with 1,000 processors or more; an example of scalability is shown in Figure 2.

4 Applications of UNRES

4.1 Folding Dynamics

UNRES/MD provides an about 4000- and 200-fold speed-up compared with all-atom MD with explicit and implicit solvent, respectively[9], thus making *ab initio* folding simulations possible. We studied[9] the folding of a number of α-helical proteins and an $\alpha + \beta$-protein (1E0G). With the version of the UNRES available at that time the simulated folding of the α-helical proteins followed the diffusion-collision model, while the folding of 1E0G started from the formation of all-α-helical structures with subsequent approach, straightening, and final packing of the ends into a two-stranded β-sheet.
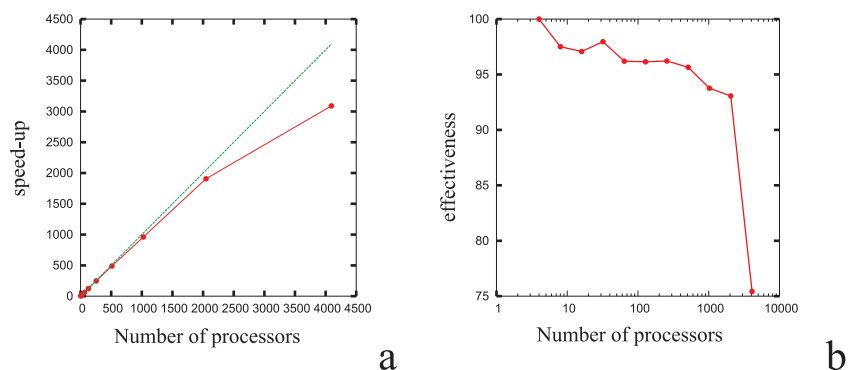
Figure 2. (a) Speedup of the UNRES/REMD algorithm as a function of the number of processors (red solid line and circles) compared with the theoretical line corresponding to 100 % speedup (green dashed line). (b) Efficiency (speedup/number of processors) vs. the number of processors for the same system. The protein studied was 1SAP proteins and the simulations were run on the Jülich Blue Gene supercomputer.

4.2 Folding Kinetics

The advantage of parallel computing combined with the speed up offered by UNRES compared to an all-atom approach enables us to run multiple Langevin trajectories simultaneously from which information about the mechanism and kinetic equations can be extracted. We carried out such a study[14] on the N-terminal fragment of the B-domain of staphylococcal protein A (a 46-residue protein with a three-helix-bundle structure). We found two folding routes: a fast one proceeding directly to the native structure and a slow one passing through an intermediate with mispacked α-helices. We also found that the variation of the fraction of native structure obeys biexponential kinetics, which is in full agreement with the presence of two folding routes.

4.3 Folding Thermodynamics

UNRES/MREMD is a robust tool to compute the thermodynamical and structural characteristics of proteins at various temperatures and, thereby, to determine the thermodynamics of protein folding. We implemented[7] the weighted histogram analysis method (WHAM)[15] method to process the results of MREMD simulations. The computed curves of heat capacity and ensemble-averaged native-likeness (e.g., RMSD from the experimental structure) are good measures of the quality of the force field[7,12].

4.4 Prediction of Protein Structure

We define the native structure as the most probable conformational ensemble at a temperature below that of the folding transition. We developed a protocol[7] in which we run UNRES/MREMD simulations, then determine the heat-capacity curve and, finally, run a cluster analysis and select the clusters with the greatest probability at a temperature below the folding-transition temperature. Figure 3 shows the results of the implementation of this protocol to predict the structure of target T0300 in the CASP7 blind-prediction test.
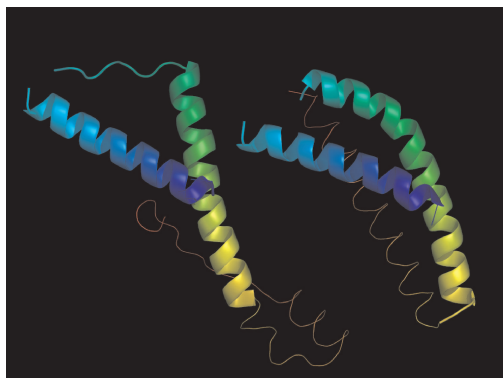
Figure 3. Comparison of the experimental (left) and predicted structure of the 2–65 fragment of the CASP7 target T0300 (RMSD=6 Å).

5   Concluding Remarks

The UNRES model has two advantages: it is a coarse-grained model with which to study protein folding in real time and is based on clear physical principles. The present UN-RES force field is still in the process of tuning but can already be used to carry out reliable prediction of protein structure, folding thermodynamics, and folding kinetics. Further applications include studying large-scale motions of proteins and protein complexes and membrane proteins.

Acknowledgments

References

1. J. Moult. Rigorous performance evaluation in protein structure modelling and implications for computational biology. *Phil. Trans. R. Soc. B*, 361:453–458, 2006.

2. H. A. Scheraga, A. Liwo, S. Ołdziej, C. Czaplewski, J. Pillardy, D. R. Ripoll, J. A. Vila, R. Kaźmierkiewicz, J. A. Saunders, Y. A. Arnautova, A. Jagielska, M. Chinchio, and M. Nanias. The protein folding problem: global optimization of force fields. *Frontiers in Bioscience*, 9:3296–3323, 2004.

3. S. Ołdziej, C. Czaplewski, A. Liwo, M. Chinchio, M. Nanias, J. A. Vila, M. Khalili, Y. A. Arnautova, A. Jagielska, M. Makowski, H. D. Schafroth, R. Kaźmierkiewicz, D. R. Ripoll, J. Pillardy, J. A. Saunders, Y.-K. Kang, K. D. Gibson, and H. A. Scheraga. Physics-based protein-structure prediction using a hierarchical protocol based on the UNRES force field – test with CASP5 and CASP6 targets. *Proc. Natl. Acad. Sci. U.S.A.*, 102:7547–7552, 2005.

4. C. B. Anfinsen. Principles that govern the folding of protein chains. *Science*, 181:223–230, 1973.

5. A. Liwo, C. Czaplewski, J. Pillardy, and H. A. Scheraga. Cumulant-based expressions for the multibody terms for the correlation between local and electrostatic interactions in the united-residue force field. *J. Chem. Phys.*, 115:2323–2347, 2001.

6. R. Kubo. Generalized cumulant expansion method. *J. Phys. Soc. Japan*, 17:1100–1120, 1962.

7. A. Liwo, M. Khalili, C. Czaplewski, S. Kalinowski, S. Ołdziej, K. Wachucik, and H.A. Scheraga. Modification and optimization of the united-residue (UNRES) potential energy function for canonical simulations. I. Temperature dependence of the effective energy function and tests of the optimization method with single training proteins. *J. Phys. Chem. B*, 111:260–285, 2007.

8. J. Lee, H. A. Scheraga, and S. Rackovsky. New optimization method for conformational energy calculations on polypeptides: Conformational space annealing. *J. Comput. Chem.*, 18:1222–1232, 1997.

9. A. Liwo, M. Khalili, and H. A. Scheraga. Molecular dynamics with the united-residue (UNRES) model of polypeptide chains; test of the approach on model proteins. *Proc. Natl. Acad. Sci. U.S.A.*, 102:2362–2367, 2005.

10. A. V. Rojas, A. Liwo, and H.A. Scheraga. Molecular dynamics with the united-residue force field: Ab initio folding simulations of multichain proteins. *J. Phys. Chem. B.*, 111:293–309, 2007.

11. M. Chinchio, C. Czaplewski, A. Liwo, S. Ołdziej, and H. A. Scheraga. Dynamic formation and breaking of disulfide bonds in molecular dynamics simulations with the UNRES force field. *J. Chem. Theor. Comput.*, 2007, submitted.

12. M. Nanias, C. Czaplewski, and H. A. Scheraga. Replica exchange and multicanonical algorithms with the coarse-grained united-residue (UNRES) force field. *J. Chem. Theor. Comput.*, 2:513–528, 2006.

13. Y. M. Rhee and V. S. Pande. Multiplexed-replica exchange molecular dynamics method for protein folding simulation. *Biophys J.*, 84:775–786, 2003.

14. M. Khalili, A. Liwo, and H. A. Scheraga. Kinetic studies of folding of the b-domain of staphylococcal protein a with molecular dynamics and a united-residue (UNRES) model of polypeptide chains. *J. Mol. Biol.*, 355:536–547, 2006.

15. S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg. The weighted histogram analysis method for free-energy calculations on biomolecules. I. the method. *J. Comput. Chem.*, 13, 1992.