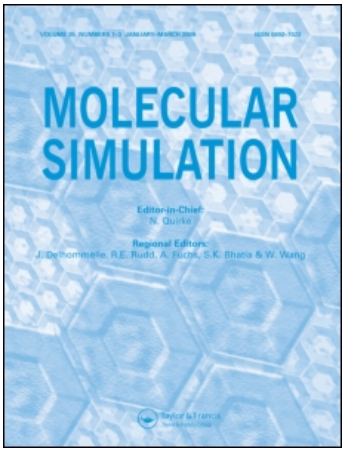

Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

Accelerating all-atom protein folding simulations through reduced dihedral barriers

R. G. Endres[a]
[a] Princeton University, Princeton, NJ 08544-1014, USA; NEC Laboratories America, Inc., 4 Independence Way, Princeton, NJ 08540, USA; Oak Ridge National Laboratory, USA

To cite this Article Endres, R. G.(2005) 'Accelerating all-atom protein folding simulations through reduced dihedral barriers', Molecular Simulation, 31: 11, 773 — 777
To link to this Article: DOI: 10.1080/08927020500266128
URL: http://dx.doi.org/10.1080/08927020500266128

Molecular Simulation, Vol. 31, No. 11, 15 September 2005, 773–777

Taylor & Francis
Taylor & Francis Group

Accelerating all-atom protein folding simulations through reduced dihedral barriers

R. G. ENDRES*

Princeton University, Princeton, NJ 08544-1014, USA; NEC Laboratories America, Inc., 4 Independence Way, Princeton, NJ 08540, USA; Oak Ridge National Laboratory, Oak Ridge, TN, 37831-6164, USA

(Received June 2005; in final form July 2005)

Protein folding requires extensive changes of backbone and sidechain dihedral angles, whose energy barriers constitute obstacles for the folding kinetics. Folding of small proteins is furthermore thought to be path-independent. Here, we propose that time-consuming all-atom protein folding simulations may be accelerated through a reduction of the dihedral barriers of the force field. In order to investigate this hypothesis, we performed various folding simulations of two small proteins. We report an acceleration towards smaller root-mean-square deviations from the native protein structure using our proposed method.

Keywords: Protein folding; Molecular dynamics; Force field; Implicit solvent model; Generalized Born model; AMBER

1. Introduction

Anfinsen's *thermodynamic hypothesis* suggests that small (monodomain) proteins fold into their native state in a path-independent way [1]. This led to the concept of viewing the rugged energy landscape [2] of protein conformations as a funnel [3,4] which guides the protein sequence to the structure of the global free-energy minimum. Realistic all-atom protein folding simulations are very informative, but suffer from being kinetically trapped in local energy minima [5–7]. Here, we investigate the new idea of smoothing the rugged energy landscape to accelerate and guide the folding simulation. Similar directed search algorithms are well known from bioinformatics [8].

Folding requires extensive changes of the backbone and sidechain dihedral angles. The energy barriers separating minima of the dihedral potential energy are larger than the available thermal energy (~ 1 kcal/mol ≈ 1.5 kT). Hence, a reduction of the barriers in the force field is expected to lead to fewer obstacles along the folding pathway (figure 1). Other sources of local minima and barriers remain: non-native but stabilizing hydrogen bonds, as well as favorable electrostatic, van-der-Waals and hydrophobic interactions. A reduction of the dihedral barriers comes also at a price; the modified force field is inaccurate. Assuming folding is path-independent, then fast initial folding with a smooth but inaccurate force field, and later refinement with the correct force field may achieve an acceleration of the simulation. This approach is related to simulated annealing, but selectively raises the temperature of the dihedral transitions initially.

We conducted various classical molecular mechanics simulations of two small proteins using an implicit solvent model. Our results show that folding can be accelerated using reduced dihedral barriers. Remaining difficulties of the folding process are traced to weaknesses of the solvent model.

2. Methods

We studied two small proteins whose NMR-structures are available: the 36-residue long villin head-piece (PDB code: 1VII), and protein-A (PDB code: 1BDD, residues 10–55). The AMBER7 molecular mechanics package with the generalized Born implicit solvent model (GB/SA) was used for all simulations [9]. There are two recently developed force fields (*parm99* and modified *parm99MOD* [10]) and two GB/SA model (switch $igb = 1$ and $igb = 2$). Since our goal is to study protein folding, for each protein we chose the combination of force field and GB/SA model, which stabilizes the folded structure during a 1 ns-long simulation starting from the folded NMR structure. The smallest root-mean-square-deviation (RMSD) of the backbone C_α atoms with respect to the NMR structure and the smallest relative RMSD fluctuations led us to choose *parm*99, $igb = 1$ for 1VII, and *parm*99*MOD*, $igb = 1$ for 1BDD.

*Corresponding author. Tel.: +609-258-8697. Fax: +609-258-8616. E-mail: rendres@molbio.princeton.edu

Molecular Simulation
ISSN 0892-7022 print/ISSN 1029-0435 online © 2005 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/08927020500266128

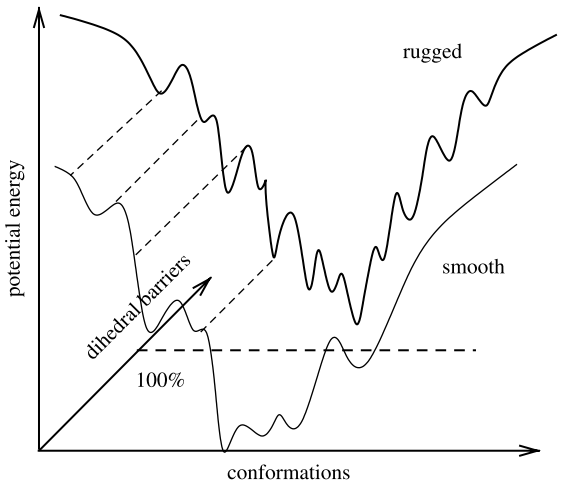

Figure 1. Schematics of protein folding funnel and its dependence on the dihedral barriers.

The accelerated folding simulations on a smoothed potential energy surface were conducted as follows. All dihedral force constants were initially reduced to 70%, and then step-wise increased to 80, 90, and 100% over a time scale of a few ns. For comparison, a conventional folding simulation (with a 100% dihedral barriers), as well as a simulation of the folded protein was performed.

2.1 Molecular dynamics simulation details

All folding simulations were started from the fully stretched amino acid sequences, which were created with the AMBER7 *xLeap* module and the *sequence* command. The simulations of the folded proteins were started from the NMR structures as obtained from the protein data bank. Initially all structures were minimized with 100 steepest descent steps, while constraining all atoms with a harmonic force with a force constant of 1.0 kcal/(molÅ^2). Subsequently, the force constant was reduced to 0.1 kcal/(molÅ^2), and the structures were heated and equilibrated at 300 K for 20 ps. The production simulations were performed without constraining any atoms; overall translational and rotational motion of the center of mass was removed every 500 steps. All molecular dynamics simulations employed the SHAKE method to fix hydrogen-heavy atom distances allowing a 2 fs integration time-step. The cut-off for long-range interactions was set to 100 Å. This simulation protocol is very similar to a recently published work [11]. On the other hand, the softened force field of the folding simulations with the reduced dihedral barriers was obtained by multiplying all dihedral force constants in the topology file by 0.7, 0.8 or 0.9. Hundred percent dihedral barriers correspond to the original topology file.

In order to analyze the simulations, we calculated the RMSD of the trajectories with the AMBER7 analysis tool *ptraj*. The backbone C_α atoms from all residues were included, the reference structure was the original NMR structure. Furthermore, the surface area was obtained from the SA energy of the standard AMBER output by dividing the SA energy by the surface tension 0.005 kcal/(molÅ^2).

2.2 Comparison of force fields and generalized born models

For each protein, one force field and one GB/SA was selected, which stabilized the folded structure. This can be regarded as a minimum requirement for successful folding. The smallest RMSD and the smallest relative RMSD fluctuations were taken as reasonable criteria. Two different force fields and two different GB/SA models were considered resulting in four potential combinations. The two force fields were the standard *parm*99 and modified *parm*99*MOD*. The *parm*99*MOD* force field is identical to *parm*99, but uses refitted Φ/ψ backbone dihedral parameters [10]. The two GB/SA models were invoked within the AMBER7 *sander* module by setting either $igb = 1$ and $PBradii = mbondi$, or $igb = 2$ and $PBradii = bondi$. Both models also required setting $gbsa = 1$ in order to calculate the surface energy (see also [11,12]). Figure 2 shows the results of the comparison over a simulation time of 1 ns. As for 1VII, the two combinations *parm*99, $igb = 1$ and *parm*99*MOD*, $igb = 1$ gave similar small RMSD values. Folding simulations using soft dihedral barriers, however, resulted in smaller RMSD for *parm*99, $igb = 1$ (data not shown). Hence, it was considered the more suitable force field. In case of 1BDD, the combination *parm*99*MOD*, $igb = 1$ was chosen, but *parm*99*MOD*, $igb = 2$ may be similarly suitable.

3. Results and discussion

Figures 3 and 4 shows the respective results of 8 ns-long simulations of 1VII and 10 ns-long simulations of 1BDD. As expected, the potential energy of the folding

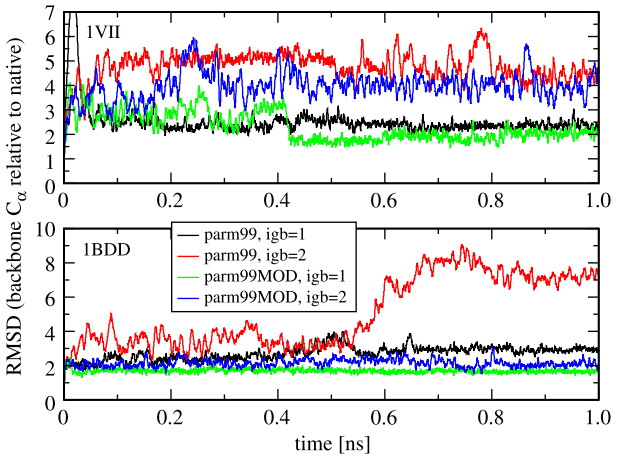

Figure 2. Comparison of two force fields (*parm*99 and *parm*99*MOD*) and two generalized Born models (AMBER7 settings $igb = 1$ and $igb = 2$) by means of 1 ns long simulations of the folded protein.

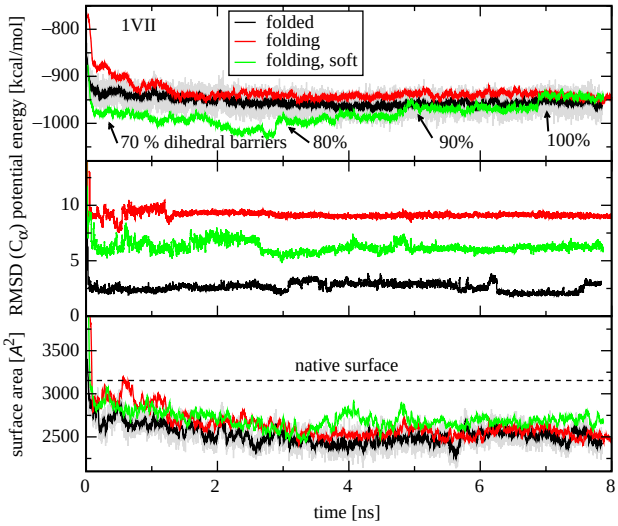

Figure 3. Representative 8 ns-long simulations of protein 1V11 using the force field *parm*99 and the solvent model *igb* = 1. Top and bottom panels show running averages over 20 ps, except for the gray background of the back curve.

simulations with the softened dihedral barriers is initially lowered, but increases later during the simulation when the accuracy of the force field is fully restored. Unfortunately, the potential energy reaches values comparable to the conventional folding simulation with 100% dihedral barriers at the end of the simulation, and not the potential energy of the targeted folded protein. The RMSD values nevertheless drop very quickly during the initial hydrophobic collapse. This results in an improvement of the RMSD values by 33% for 1VII and 25% for 1BDD compared to the conventional folding simulations. Surprisingly, all folding simulations do not improve further after the first few ps. This may indicate problems with the force field or the solvent model.

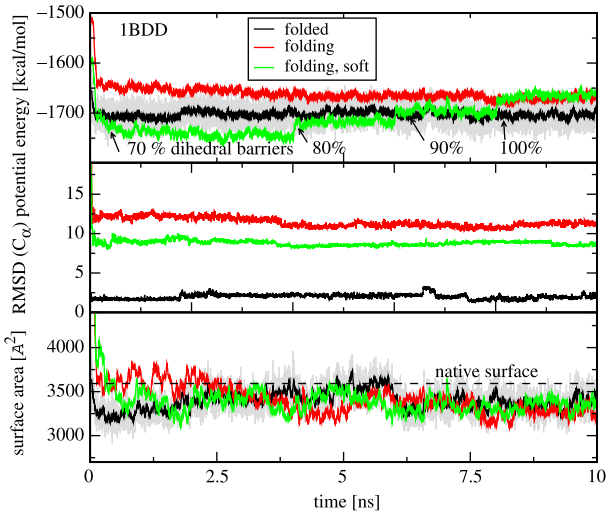

Figure 4. Representative 10 ns-long simulations of protein 1BDD using the force field *parm*99 *MOD* and the solvent model *igb* = 1.

3.1 Deficiency of the solvent model

Small proteins fold on a μs time scale [4] as observed in *in vitro* experiments, but the reduced viscosity of the implicit solvent model is expected to reduce folding times drastically—possibly to a few ns [11]. However, one possible reason for generally slow and trapped folding simulations may be inferred by plotting the surface area along the trajectory. As shown in bottom panel of figures 3 and 4, all surface areas quickly shrink below the surface area of the NMR structure (dotted line). This bias towards compact structures may stem from the GB/SA implicit solvent model. The GB dielectric continuum model provides a good analytical approximation of numerical solutions of the Poisson-Boltzmann equation. However, the SA model applies, indiscriminate of the atom-type, a surface tension of 0.005 kcal/(mol$\mathring{A}^2$) to the total solvent-accessible surface area in order to model the hydrophobic effect. The GB/SA model was originally designed for hydrocarbons and small organic molecules [13], and is clearly insufficient for the large variety of polar and charged amino acids. Applying the surface tension mainly to solvent-exposed carbons may constitute a potential improvement, and may lead to less compact and easier-to-fold intermediate structures. A similar atom-type dependent philosophy is well-known from the atomic-solvation parameters used as an alternative means for estimating solvation free energies [14].

3.2 Dependence of folding on initial conditions

In the following, the robustness of the behavior of the folding simulations is addressed. According to the simulations shown in figures 3 and 4, the main improvement of the softened force field occurs during the first ns. We performed two additional 1 ns-long folding simulations for each protein (1VII and 1BDD), for both the unmodified force field and the softened force field with 70% dihedral barriers. The different initial conditions were created through variation of the equilibration time by a few ps. The results are shown in figure 5, which also contains the simulations from figures 3 and 4 (black, red, and green curves). In case of 1VII, the conventional folding simulations with the unmodified force field are all very similar and have large RMSD values. While two simulations using the softened potential show quick improvement toward smaller RMSD values, the third (the brown curve) initially improves, but then approaches the conventional folding simulations. As for 1BDD, all three simulations with the softened force field and even two with the unmodified force field show comparable small RMSD values. From this we conclude that the softened force field leads to smaller RMSD values on average.

3.3 Effect of dihedral barriers on kinetics and structure

Although the dihedral barriers constitute mainly obstacles for the folding kinetics, they nevertheless are somewhat

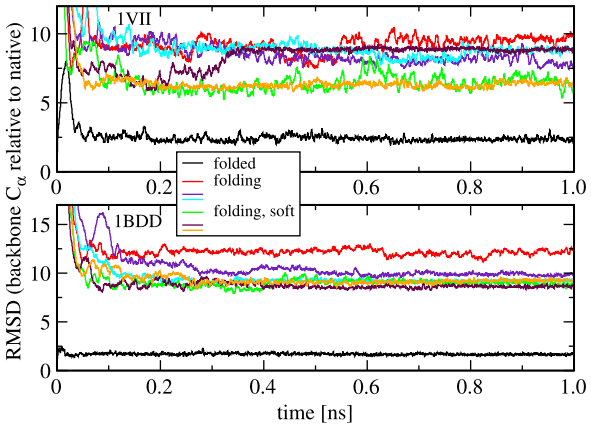

Figure 5. Dependence of the folding simulations on initial conditions. Force fields: 1VII—parm99, *igb* = 1; 1BDD—*parm*99*MOD*, *igb* = 1; black: folded; red, purple, and light blue: folding; green, brown, and orange: folding with 70% dihedral barriers.

important for the intermediate and native protein structures. This can be seen qualitatively as follows. The dihedral potential around a bond of sp^3-hybridized atoms of the backbone or sidechains has generally three minima; two are located around $\pm 60°$ and one is near 180°. The distribution of backbone ϕ and ψ dihedral angles of experimentally determined protein structures can be seen in Ramachandran plots [15]. α-helices have ψ angles close to $-60°$, but their ϕ angles have values between $-180°$ and $-60°$. β-sheets have very broad distributions. Their ϕ and ψ angles are mostly located in the intervals between $-180°$ and $-60°$ and between 60° and 180°, respectively. As for loops, both angles are relatively sharply peaked around 60°. Hence, dihedral angles do not only assume the value of minimal dihedral energy, but instead may deviate in order to lower the van-der-Waals interactions, the electrostatic interactions, the hydrogen bonds, or the solvation free energy. This trade-off may lead to a net lowering of the total free energy. Hence, a reduction of the barriers is expected to affect the protein structure, which makes it necessary to restore the accurate force field toward the end of the simulation.

4. Conclusions

In this study, we describe an attempt to accelerate all-atom protein folding simulations by simply reducing the dihedral barriers of the force field at the beginning of the simulation. This contrasts recently proposed mathematically more rigorous methods. Among the most promising schemes to enhance conformational sampling on rough energy surfaces are molecular dynamics algorithms which use nonlinear variable transformations to reduce energy barriers [16] and allow longer integration time steps (100 fs) [17], as well as flat-histogram Monte Carlo algorithms which perform random walk in energy space independent of temperature [18]. These algorithms still have to prove their capabilities with respect to folding of realistic atomistic proteins. For our present study we first tested different parameterizations of force fields and generalized Born implicit solvent models for their suitability, and then performed various 8–10 ns-long simulations of two small proteins, 1VII and 1BDD. While our method leads to somewhat smaller root-mean-square deviations from the native structure as compared to the unmodified force field, the folding simulations get trapped for both the modified and unmodified force fields. In order to find the likely origin of this problem, we monitored the protein surface area along the trajectories. Our analysis indicates that the application of the surface tension indiscriminate of the atom type leads to a bias toward compact intermediate structures, which are hard to fold further. Yet another difficulty may exist in the way we evaluate the folding progress of our test proteins. For instance, it is not clear whether the classical force fields used here even permit folding paths to the experimental native structures. Although the folded protein structures used in this study are relatively stable and have a lower energy than the initial unfolded proteins, they may be inaccessible on the potential energy surface or not the true groundstates. Even shorter peptides, which knowingly fold, could be used instead.

Acknowledgements

The author is grateful to Alexey Onufriev and Ned Wingreen for helpful discussions. This work was supported by DOE under Contract DE-AC05-00OR22725 with UT-Battelle LLC.

References

[1] C. Anfinsen. Principles that govern folding of protein chains. *Science*, **181**, 223 (1973).
[2] H. Frauenfelder, S.G. Sligar, P.G. Wolynes. The energy landscapes and motions of proteins. *Science*, **254**, 1598 (1991).
[3] P.G. Wolynes, J.N. Onuchic, D. Thirumalai. Navigating the folding routes. *Science*, **267**, 1619 (1995).
[4] J. Schonbrun, K.A. Dill. Fast protein folding kinetics. *Proc. Natl. Acad. Sci. USA*, **100**, 12678 (2003).
[5] U.H.E. Hansmann. Protein folding *in silico*: an overview. *Comput. Sci. Eng.*, **5**, 64 (2003).
[6] S. Gnanakaran, H. Nymeyer, J. Portman, K.Y. Sanbonmatsu, A.E. Garcia. Peptide folding simulations. *Curr. Opin. Struct. Biol.*, **13**, 168 (2003).
[7] J.N. Onuchic, P.G. Wolynes. Theory of protein folding. *Curr. Opin. Struct. Biol.*, **14**, 70 (2004).
[8] J.U. Bowie, D. Eisenberg. An evolutionary approach to folding small alpha-helical proteins that uses sequence information and an empirical guiding fitness function. *Proc. Natl. Acad. Sci. USA*, **91**, 4436 (1994).
[9] D.A. Case, D.A. Pearlman, J.W. Caldwell, T.E. Cheatham III, J. Wang, W.S. Ross, C.L. Simmerling, T.A. Darden, K.M. Merz, U.C. Singh, P.K. Weiner, P.A. Kollman. *AMBER 7*, San Francisco, University of California (2002).
[10] C. Simmerling, B. Strockbine, A.E. Roitberg. All-atom structure prediction and folding simulations of a stable protein. *J. Am. Chem. Soc.*, **124**, 11258 (2002).
[11] A. Onufriev, D. Bashford, D.A. Case. Exploring protein native states and large-scale conformational changes with a modified

generalized Born model. *Proteins: Struct. Funct. Bioinform.*, **55**, 383 (2004).

[12] A. Onufriev, D. Bashford, D.A. Case. Modification of the generalized Born model suitable for macromolecules. *J. Phys. Chem. B*, **104**, 3712 (2000).

[13] W.C. Still, A. Tempczyk, R.C. Hawley, T. Hendrickson. Conformations of cycloheptadecane—a comparison of methods for conformational searching. *J. Am. Chem. Soc.*, **112**, 6127 (1990).

[14] L. Wessen, D. Eisenberg. Atomic solvation parameters applied to molecular-dynamics of proteins in solution. *Prot. Sci.*, **1**, 227 (1992).

[15] C. Branden, J. Tooze, *Introduction to Protein Structure*, Garland Publishing, New York (1999), Chapter 1.

[16] Z. Zhu, M.E. Tuckerman, S.O. Samuelson, G.L. Martyna. Using novel variable transformations to enhance conformational sampling in molecular dynamics. *Phys. Rev. Lett.*, **88**, 100201 (2002).

[17] P. Minary, M.E. Tuckerman, G.J. Martyna. Long time molecular dynamics for enhanced conformational sampling in biomolecular systems. *Phys. Rev. Lett.*, **93**, 150201 (2004).

[18] N. Rathore, J.J. De Pablo. Monte Carlo simulations of protein through random walk in energy space. *J. Chem. Phys.*, **116**, 7225 (2002).