PROTEINS: Structure, Function, and Genetics 49:439–445 (2002)

SHORT COMMUNICATION

All-Atom Fast Protein Folding Simulations: The Villin Headpiece

Min-yi Shen and Karl F. Freed*
James Franck Institute and Department of Chemistry, The University of Chicago, Chicago, Illinois

ABSTRACT **We provide a fast folding simulation using an all-atom solute, implicit solvent method that eliminates the need for treating solvent degrees of freedom. The folding simulations for the 36-residue villin headpiece exhibit close correspondence with the landmark all-atom explicit solvent molecular dynamics simulations by Duan and Kollman (Duan & Kollman, Science 1998;282:740–744; Duan, Wang, & Kollman, Proc Natl Acad Sci USA 1998;95:9897–9902). Our implicit solvent approach uses only an entry-level single CPU PC with comparable throughput (~4 nsec/day) to the DK supercomputer simulation. The native state is shown to be stable. Our 200-nsec folding trajectory agrees with the DK simulation in displaying a burst phase, a rapid initial shrinkage, a highly native-like binding site structure, and more. Proteins 2002;49:439–445.**

An understanding of the physical process governing protein folding provides the missing link between the primary amino acid sequence and the biological function of proteins. One main impediment to describing protein folding dynamics in solution is the need for explicitly including solvent degrees of freedom, which generally constitute more than 95% of the atoms in the system. This description of the protein folding dynamics is currently just beyond computer capabilities, so faster, accurate methods for exploring the folding energy landscape are essential but remain one of the great challenges for modern computational biophysics. The introduction of an implicit solvent model, in which the solvation effect is considered in a mean-field representation, can dramatically increase the throughput of simulations and still retain reasonable accuracy. In this study, we introduce a simple and fast implicit solvent model that can accurately describe the dynamics of the protein folding and that can be extended to systems containing other solvents. The application of this powerful method is best illustrated by considering recent state-of-the-art molecular dynamics simulations for the 36-residue villin headpiece (HP-36).

The chicken villin HP-36[3–5] in aqueous solution displays a stable native structure with full F-actin binding function.[6–8] Its folding time is estimated as 1–10 μsec by solvation free energy calculations,[9] making it one of the fastest folding proteins and thus natural for theoretical treatment. The landmark all-atom MD simulations for HP-36 by Duan and Kollman (DK)[1] exhibit hydrophobic collapse as the main driving force in the initial stages of protein folding, followed by a slower structural adjustment process. DK's explicit water simulation uses 256 CPUs and highly proprietary codes for the Cray T3D/E to produce 5 nsec of trajectory per day, corresponding to more than one single CPU year for a 10-nsec trajectory. Hence, the ideal implicit solvent description must dispense with the costly explicit treatment of solvent coordinates but must accurately describe the solvent effects in providing (1) friction sources and random forces, (2) free energy of solvation, and (3) a dielectric medium to screen electrostatic interactions from solvent particles. Here we purpose a fast implicit solvent model that contains all three necessary elements stated above (see Methods).

Many implicit solvent potential functions have been considered before.[10–18] However, only few contain all three necessary elements of solvation, and none has been subjected to the most stringent test of whether the reduced model can successfully describe the dynamics of the folding of a random-coil chain to a native-like structure in accord with the all-atom detail generated by MD simulations that include the solvent. Among the widely used implicit solvent models[17,18] are the generalized Born/surface area[10] (GB/SA) and the effective energy function 1[16] (EEF1) approaches. The GB method provides a very detailed (and more accurate) treatment for the electrostatic contributions (at the cost of increased computer time by roughly a factor of two to three compared with our model in tests for Met-enkephalin). However, tests of the GB/SA model in Langevin Dynamics (LD) simulations for Met-enkephalin exhibit mediocre agreement with the corresponding explicit water simulations[19]; the GB/SA model

Grant sponsor: National Institutes of Health; Grant number: GM56678.

*Correspondence to: Karl F. Freed, James Franck Institute and Department of Chemistry, The University of Chicago, IL 60637. E-mail: k-freed@uchicago.edu

Received 19 January 2002; Accepted 12 April 2002

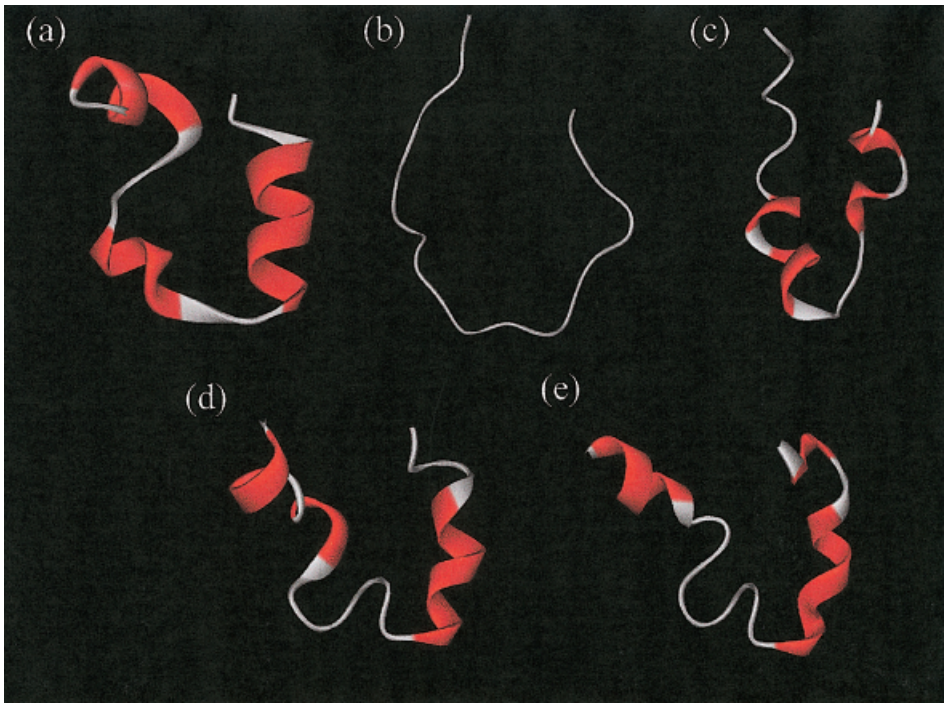

Fig. 1. Snapshots of structures in (**a**) native state, (**b**) denatured state, and at (**c**) 14, (**d**) 59, and (**e**) 164 nsec in the folding simulation. Red and white indicate helix and coil, respectively. First residue Met1 is always on the left. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

represents Met-enkephalin as too compact. The EEF1 approach introduces a sophisticated Gaussian model for the solvation energy but only uses a simple $\epsilon(r) = r$ distance dependent dielectric constant, which is less than satisfactory.[†] As demonstrated by our extensive tests for Met-enkephalin[19] and by the computations described here for the villin HP-36, our implicit water model yields a simple and rapid method that treats all three solvation effects in a balanced manner while still retaining good accuracy in comparison with explicit water MD simulations.

In more detail, we have tested our implicit solvent LD description of solvated protein dynamics for the small, flexible neurotransmitter peptide Met-enkephalin. This initial test has been chosen because the solvated Met-enkephalin does not exhibit a single native structure but instead shuttles between extended, semi-packed, and packed conformations with quite distinct overall radii of gyration, so the peptide ranges through a wide variety of conformations. In addition to the 200-fold speed up over the explicit water MD simulation that we obtain from the LD simulations,[19] there is good agreement between explicit water MD and implicit water LD simulations for a variety of equilibrium and dynamical properties of the Met-enkephalin. For example, the equilibrium distribution of the radius of gyration for Met-enkephalin agrees well with the MD simulations but exhibits our implicit solvent model as slightly too hydrophobic. Equilibrium conformational populations for the dihedral angles about tyrosine are in excellent agreement with MD simulations, while the conformational distributions for the phenylalanine slightly favor gauche conformations (again because the model is slightly too hydrophobic). Numerous time-correlation functions for both local and global motions of Met-enkephalin likewise display good agreement between explicit and implicit water simulations.

The moderate size of the villin HP-36 enables us to continue and extend our tests of the implicit solvent description using only a single personal computer (a mildly equipped Athlon 800 MHz/128 Mb) and a time duration of 1–2 months. A slight extension of the prior implicit solvent treatment (see Methods) is necessary for treating the charged and polar residues that are present in the villin HP-36. The main goals are to determine whether the implicit LD dynamics represent a good approximation to the MD simulations of DK, and, if so, to employ the LD simulations as a vehicle for probing the atomistic scale dynamics of protein folding.

Figure 1(a) displays the tri-helical native structure of HP-36. Following the identical procedure used by DK, our test of the implicit water model begins with a 50-ns equilibrium simulation at 300K that starts with the NMR structure.[5] Figure 2(a) displays the radius of gyration, and the segmental RMSD for the equilibrium simulation. The radius of gyration (R_g) jumps at the very early stages (0 to 5 nsec) to a maximum of 11.58 Å at 5.01 nsec. Then, R_g

[†]We are unable to test the speed and accuracy of the EEF1 method because it is a built-in feature of CHARMM which only allows the usage of CHARMM force fields.

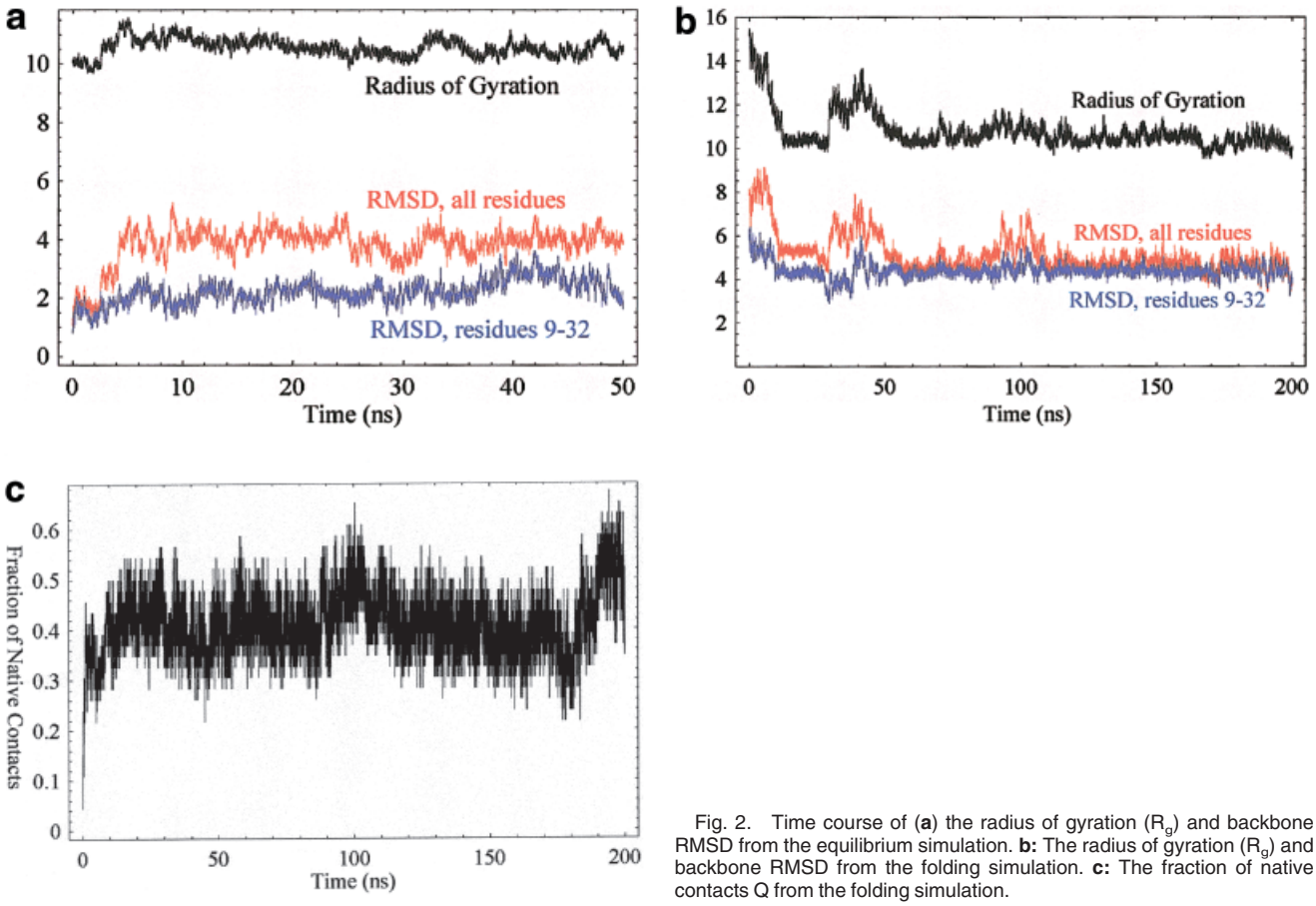

Fig. 2. Time course of (**a**) the radius of gyration (R_g) and backbone RMSD from the equilibrium simulation. **b:** The radius of gyration (R_g) and backbone RMSD from the folding simulation. **c:** The fraction of native contacts Q from the folding simulation.

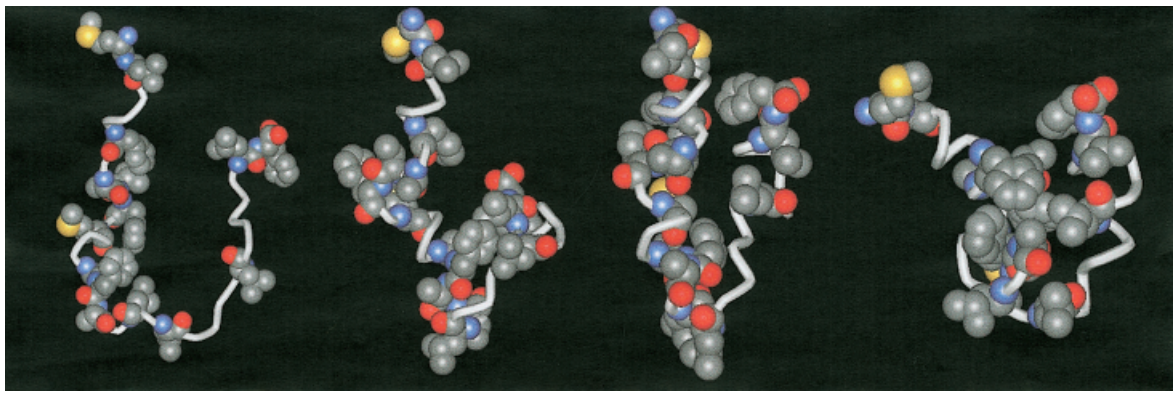

Fig. 3. Distribution of hydrophobic residues in different time frames. First residue Met1 is always placed on the left. From left to right: denatured, 8, 14, and 65 nsec in the folding simulation.

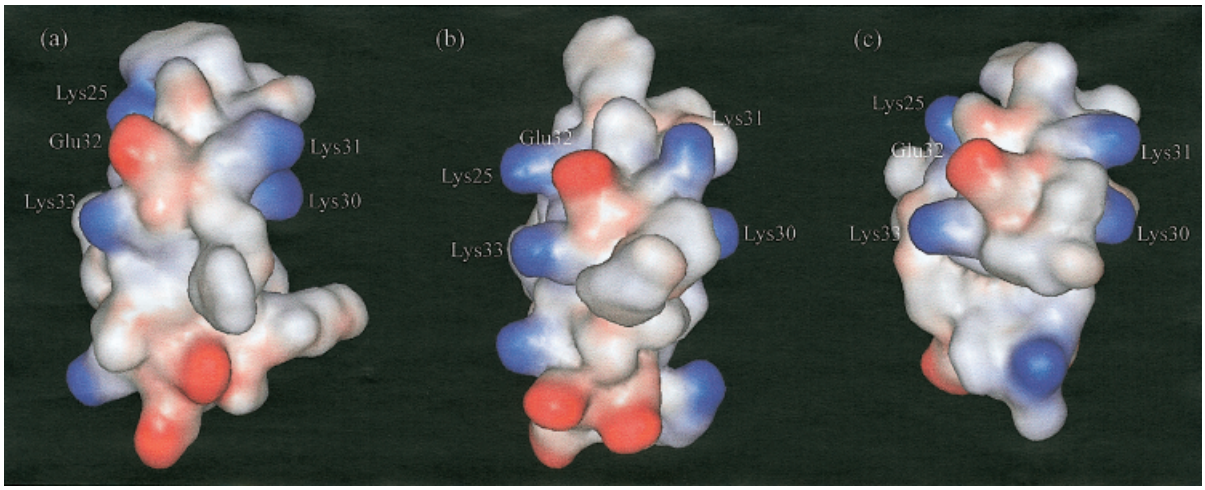

Fig. 4. Electrostatic potential surfaces for the F-actin binding site of (**a**) native state, (**b**) at 30 nsec from stability simulation, and (**c**) at 196 nsec from folding simulation. Blue and red indicate positive and negative electrostatic potentials, respectively.

shrinks continuously towards its minimum (9.96 Å) at 30.50 nsec. After this structural adjustment stage, R_g remains stable throughout the whole 50-nsec trajectory, with fluctuations in R_g mainly due to motions of the N-terminus leading portions. Residues 1–8 display larger scale motions and a wider RMSD distribution (from the NMR structure) than the central 9–32 residues. This feature agrees well with DK's controlled simulation that uses the same force field and the TIP3P model for water.[20] The overall RMSD also jumps in the early stages, but then stabilizes at ~4 Å after 5 nsec. The RMSD for the central residues 9–32 is much more stable and only fluctuates near 2 Å for the later stages of the equilibrium simulation, a result again agreeing quite well with DK's 1.5 Å RMSD. After the 50-ns equilibrium simulation, most of the hydrophobic core is preserved, and only the terminal residues undergo larger scale motions: the Phe36 rotates toward the first helix to form a hydrophobic core, and Met1 and Leu2 exit the core. Except for these minor changes in the terminal residues, the hydrophobic core remains intact during the whole simulation, verifying the adequacy of our implicit solvent approach, confirming the stability of the experimental NMR structure, and exhibiting similar behavior to the DK stability simulations.

Again, following the identical denaturing scheme used by DK, Figure 1(b) displays the HP-36 structure after a 1-nsec denaturing simulation at 1,000 K. After heating, the HP-36 structure is devoid of helical content (or any secondary structure) and only preserves ~6.5% of the backbone native contacts. Although the U-shaped backbone structure remains, the source of this curved shape in the native state, the hydrophobic core, is totally destroyed. While this backbone structural feature might be eliminated with longer heating, this procedure is not considered because one main purpose of the present study is to compare our implicit water treatment with the explicit water simulation by DK, so the same simulation protocols are followed. Moreover, this denaturing protocol is probably more drastic than is likely in actual experiments and, therefore, renders the room temperature folding more difficult than experimentally. R_g for the denatured structure is 15.44 Å, which indicates a much more extended structure than in the native state (9.44 Å). An implicit solvent LD folding simulation then proceeds from this denatured structure at a constant temperature of 300 K. Figure 2(b) displays the time evolution of R_g and the structural RMSD for two structural fragments. R_g undergoes a rapid shrinkage at the early folding stages and temporarily pauses at 10 nsec. During this initial folding phase, one arm of the U-shaped structure rapidly collapses into the central portion, leading to a sudden growth of the RMSD. A sporadic formation of α-helices occurs during this stage, but none is stable enough to survive. A second expansion process abruptly begins at 29 nsec, where R_g reaches a maximum of 13.67 Å (89% of R_g in the unfolded state) at 41.37 nsec, before settling down at 50 nsec. These two major sets of conformation adjustments are directly related to the formation of the hydrophobic core.

Figure 3 shows the time evolution of the distribution of hydrophobic residues. The rapid, early collapse of the C-terminal arm brings Leu35 and Phe36 together with Leu21 and Pro22 to form an unstable cluster. The contraction of the N-terminal arm generates a marginally stable conformation that persists for nearly 20 nsec. This intermediate structure contains two hydrophobic clusters instead of the single one present earlier. The hydrophobic Phe36-Leu35-Leu29 residues are closely packed on one side, and all other non-polar residues lie on the other side (presenting a large solvent-accessible area), resulting in a V-shaped backbone with a central "hydrophobic vacancy" in which four hydrophilic and polar residues are unfavorably trapped. The highly unfavorable side chain packing of polar and hydrophilic residues in the center of hydrophobic core results in the subsequent structural expansion process at 29 nsec. After this readjustment, a loose native-like hydrophobic core is partly formed at 40 nsec, then tightens in the next 10 nsec, and eventually settles at 50–60 nsec. The R_g and the RMSD reach a quasi-steady state between 60 and 200 nsec. This "burst phase" time agrees well with the explicit water simulation (~60 nsec). Several explicit water simulations by DK with different initial conditions all produce a similar burst phase time of 60 nsec,[21] which may imply that the folding dynamics in the early stage are not very sensitive to the starting structures. The similarity between explicit and implicit water simulations suggests the similarity of the free energy landscapes for both descriptions. The burst phase in the LD simulation displays a two-step behavior instead of the one-step burst in the explicit water simulation. This difference may be explained by the need in the explicit water simulations to expel solvent molecules during the collapse of the hydrophobic core. Because the burst phase contains two processes, a structural contraction and the formation of a hydrophobic core, the absence of explicit water molecules would speed the first contraction phase, but the time scale for the second process remains unchanged.

The lowest RMSD for all residues in the 200-nsec LD simulation is 3.46 Å, significantly lower than the minimum 4.8 Å found by DK during the first 200 nsec. The RMSD for all residues in our simulation contains larger scale fluctuations beginning at 30 and 90 nsec, marking the onsets of structural readjustment. A third hydrophobic core readjustment event in our simulation begins at 170 nsec but is not evident from the evolution of the RMSD. The time interval between these structural re-adjustments in our simulations agrees well with the contraction-expansion interval of 80 nsec in the trajectory of DK. Our LD simulation yields smaller short time scale structural fluctuations than the DK simulations where some water might unfavorably be packed in the core. Therefore, larger scale structural rearrangement is required to expel these water molecules in the MD simulations than in the implicit water approach. In addition, because newly formed hydrogen bonds stabilize the structure in the early folding stages, hydrogen bonds on the surface of the hydrophobic core are not protected from solvent attack and exhibit

different dynamics in the explicit and implicit water simulations. In the former case, unprotected hydrogen bonds are readily attacked and broken by explicit water molecules, leading to greater structural fluctuations than in the implicit water simulations where unprotected hydrogen bonds are more stable. A second folding simulation[22] starting from the midpoint (~150 nsec) of the first simulation has been completed and exhibits larger scale open-close motions at approximately 170 nsec, similar to those occurring in the DK simulation during the same time interval.

Figure 2(c) displays the time evolution for the fraction of native contacts Q for our folding simulation. Starting from 0.065 in the denatured state, Q rapidly grows to 0.46 at 1.26 nsec, drops back to 0.21, and then steadily reaches a plateau ~0.4 in the 15–20-nsec range. This plateau time for Q accords well with DK's 20 nsec. The plateau values of Q differ (0.3 for DK vs. our 0.4) presumably because of our lower RMSD. After the initial burst stage, Q reaches a maximum for two times, 0.61 at 102 nsec and 0.64 at 194 nsec, which correspond to later stages of the structural readjustments. Hence, during these readjustment stages, Q evolves toward the native state. In contrast, Q from the LD stability simulation usually fluctuates within 0.75 to 0.7, significantly higher than the average from the folding trajectory, which has not yet reached the free energy basin containing the native state. The latter behavior is expected because the estimated folding time of HP-36 is 5–50 times longer than 200 nsec. Many stable secondary structures are partly formed in the later stages of this folding simulation. Figure 1(e) displays the structural snapshot at 164 nsec, where a completely formed third helix and a skew first helix are already present. The rapid formation of the third helix is very important for villin's biological function because of the geometric requirement on the actin binding sites.

Friederich et al.[23] argued that the F-actin binding affinity of the villin headpieces is controlled by the KKEK (Lys30-Lys31-Glu32-Lys33) motif that lies near the C-terminal because mutations in this motif reduce villin's binding activity. In the native structure, the binding surface of HP-36 is located on the outer side of the third helix, and the KKEK motif forms a charged ring outside the molecular surface, with the GLF hydrophobic residues attached nearby. Close to GLF lies the polar tip of Lys25, which is believed to participate in F-actin binding. Figure 4 presents the electrostatic surface representation of the binding site. Very similar structural features are evident both in the native and in our simulated structures. Our stability simulation preserves the KKEK charged ring structure and the proximal Lys25. The Lys25 residue is nearly buried in the native structure but protrudes more in our simulation because of the solvation energy. The KKEK ring structure is well established after 196 nsec of the folding simulation, with the Lys25 slightly more distal than in the native structure.

This study displays the accuracy and rapidity of the implicit water model. Using one CPU, the folding mechanism is revealed with atomic detail in strong accord with the landmark explicit water MD simulations. The accuracy of our approach is exemplified by the all-residue RMSD prediction of 3–4 Å, and the active site geometry is correctly generated. Our method affords great flexibility in selecting computing hardware to describe protein folding and structural genomics. The accuracy and enhanced speed of the method suggest that it will also be valuable in many applications, e.g., for refining NMR solution structures and for validating folded structures deduced from more coarse-grained theories that do not preserve atomistic details.

Ultimately, we desire to describe and predict the dynamics observed in real experiments, but, unfortunately the exact interactions are unknown. Thus, implicit solvent methods can be validated by the degree to which they reproduce the dynamics provided by explicit solvent MD simulations that employ the best available all-atom interaction potentials. However, due to the enormous computational labor involved in explicit solvent MD simulations of proteins for interesting time scales ($\sim 10^8$–10^9 energy evaluations for a microsecond level simulation), this comparison cannot be made for a large number of systems in the manner, for instance, used to determine solvation potential parameters from the solvation energies for a wide variety of organic molecules. Instead, the tests of the methods for treating dynamics must be a continuing process, one that has begun with the treatment of the dynamics of a small flexible peptide, Met-enkephalin, with no native structure. The present work extends the analysis to the initial stages of protein folding where the implicit solvent model provides a remarkably similar atom-scale folding pathway as in the explicit solvent MD simulations. This close correspondence supports the view that hydrophobic collapse, followed by a series of structural rearrangements to expel trapped water and polar residues, represents the basic molecular mechanism in the initial stages of villin HP-36 folding. Future work will have to proceed to more complex processes and, undoubtedly, improve the implicit solvent models. The prior work for Met-enkephalin indicates some inaccuracies of the solvation potential stemming from their being slightly too hydrophobic in comparison with the AMBER/TIP3P potentials used in the MD simulations, a limitation that might be overcome with improved parameterization. The present work notes other limitations associated with rapid dynamics involving unprotected hydrogen bonds and the expelling of water from hydrophobic cores, limitations that are more difficult to overcome within implicit solvent models.

METHODS

The molecular simulations are performed with a modified version of the TINKER 3.7 molecular design package[24] and with the AMBER95 (parm94) force field.[25] The Ooi–Scheraga solvent-accessible surface area[13] solvation potential is added to calculate the total energy. A sigmoidal type distance dependent dielectric coefficient,[26]

$$\varepsilon(r) = D - \frac{D-1}{2}(S^2r^2 + 2Sr + 2)e^{-Sr}$$

(with $D = 78.0$, $S = 0.3$) is used for mimicking the differences between the electrostatic screening inside the protein and in bulk water. The sigmoidal type dielectric screening function has been introduced first by Mehler and Eichele,[27] and has been applied in Monte Carlo calculations[28] and folding-misfolding analyses for three standard decoy sets.[29] This sigmoidal dielectric model may be viewed as a "one size fits all" approximation to the GB model in which all atoms are treated with identical Born radii. Non-bonding interactions are truncated at 8.0 Å (a test trajectory running without a cut-off displays similar statistics).[22] The numerical integration of the Langevin equation employs the velocity Verlet algorithm[30] with step sizes of 2.0 fsec for 65 nsec of the folding trajectory and 2.5 fsec for the remainder. We use the Pastor–Karplus[31] scheme for calculating the atomic friction coefficients along with the experimental water viscosity $\epsilon = 0.84$ cp. The importance of using the sigmodal dielectric function is illustrated by a 50-nsec simulation with constant bulk water dielectric constant that fails to preserve the native structure and that finally produces a loose globular structure without helices. The simulation temperature is controlled at 300K with a Berendsen-type[32] thermal bath coupling. All trajectories are stored at 10-psec time intervals. The root-mean-square deviation (RMSD) between structures is calculated by matching the backbone C_α point sets by rigid translations and rotations using the conjugated gradient minimization algorithm.[33] The secondary structures in Figure 1 are assigned by the method suggested by Kabsch and Sander using the program DSSP.[34] The electrostatic surface plot is generated using the MSI WebLab ViewerLite 4.0. The electrostatic potential is presented in a solvent accessible fashion, where the electrostatic potential is calculated using the Gasteiger atomic charge scheme.[35] The method for estimating the speed-up gained by the implicit solvent method is explained in our previous work,[19] where the dependence of the computational cost on N (the degree of polymerization) is linear for the implicit water simulation but grows as $N^{1.5}$ in the explicit water case for extended flexible polymer chains. We obtained a 200-fold speed-up for a 5-residue chain, so a 540-fold speed-up is expected for a non-globular 36-residue peptide using the $N^{1.5}$ scaling. When the same arguments are applied to estimate the total number of atoms required for the HP-36 explicit water simulation, we obtain ~12,000 atoms, which is close to the ~10,000 atoms needed,[1] thereby supporting our estimation.

NOTE ADDED IN PROOF

Since submission of this manuscript, the implicit solvent LD simulation programs have been improved to run four times faster.[22]

ACKNOWLEDGMENTS

The authors would like to thank Steve Berry, Joan-Emma Shea and Bianxiao Cui for helpful discussions.

REFERENCES

1. Duan Y, Kollman PA. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. Science 1998;282:740–744.
2. Duan Y, Wang L, Kollman PA. The early stage of folding of villin headpiece subdomain observed in a 200-nanosecond fully solvated molecular dynamics simulation. Proc Natl Acad Sci USA 1998;95: 9897–9902.
3. Brescher A, Weber K. Villin -major microfilament-associated protein of the intestinal microvillus. Proc Natl Acad Sci USA 1979;76:2321–2325.
4. McKnight CJ, Doering DS, Matsudaira PT, Kim PS. A thermostable 35-residue subdomain within villin headpiece. J Mol Biol 1996;260:126–134.
5. McKnight CJ, Matsudaira PT, Kim PS. NMR structure of the 35-residue villin headpiece subdomain. Nature Struct Biol 1997;4: 180–184.
6. Pope B, Way M, Matsudaira PT, Weeds A. Characterization of the F-actin binding domains of villin - classification of F-actin binding-proteins into 2 groups according to their binding-sites on actin. Febs Lett 1994;336:58–62.
7. Glenny JRJ, Geisler N, Kaulfus P, Weber K. Demonstration of at least 2 different actin-binding sites in villin, a calcium-regulated modulator of F-actin organization. J Biol Chem 1981;256:8156–8161.
8. Glenny JRJ, Kaulfus P, Matsudaira PT, Weber K. F-actin binding and bundling properties of fimbrin, a major cytoskeletal protein of microvillus core filaments. J Biol Chem 1981;256:9283–9288.
9. Lee M, Duan Y, Kollman PA. Use of MM-PB/SA in estimating the free energies of proteins: application to native, intermediates, and unfolded villin headpiece. Proteins 2000;39:309–316.
10. Schaefer M, Karplus M. A comprehensive analytical treatment of continuum electrostatics. J Phys Chem 1996;100:1578–1599.
11. Still WC, Tempczyk A, Hawley RC, Hendrickson T. Semianalytical treatment of solvation for molecular mechanics and dynamics. J Am Chem Soc 1990;112:6127–6129.
12. Eisenberg D, McLachlan AD. Solvation energy in protein folding and binding. Nature 1986;319:199–203.
13. Ooi T, Oobatake M, Nemethy G, Scheraga HA. Accessible surface areas as a measure of the thermodynamic parameters of hydration of peptides. Proc Natl Acad Sci USA 1987;84:3086–3090.
14. Colonna-Cesari F, Sander C. Excluded volume approximation to protein-solvent interaction - the solvent contact model. Biophys J 1990;57:1103–1107.
15. Lazaridis T, Karplus M. "New view" of protein folding reconciled with the old through multiple unfolding simulations. Science 1997;278:1928–1931.
16. Lazaridis T, Karplus M. Effective energy function for proteins in solution. Proteins 1999;35:133–152.
17. Roux B, Simonson T. Implicit solvent models. Biophys Chem 1999;78:1–20
18. Cramer CJ, Truhlar DG. Implicit solvation models: equilibria, structure, spectra, and dynamics. Chem Rev 1999;99:2161–2200.
19. Shen MY and Freed KF. Long time dynamics of met-enkephalin: comparison of explicit and implicit solvent models. Biophys J 2002;82:1791–1808.
20. Jorgensen WL, Chandrasekhar J, Madura JJD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79:926–935.
21. Duan Y, Kollman PA. Computational protein folding: from lattice to all-atom. IBM Syst J 2001;40:297–309.
22. Shen MY, Freed KF. Unpublished results.
23. Friederich E, van Compernolle K, Huet C, Goethals M, Finidori J, Vandekerckhove J, Louvard D. An actin-binding site containing a conserved motif of charged amino-acid-residues is essential for the morphogenic effect of villin. Cell 1992;70:81–92.
24. Ponder JW, Rubenstein S, Kundrot C, Huston S, Dudek M., Kong Y, Hart R, Hodsdon M, Pappu R, Mooij W, Loeffler G. TINKER: Software tools for molecular design, version 3.7, Washington University, St. Louis; 1999
25. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM Jr., Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J Am Chem Soc 1995;117: 15179–15197.

26. Ramstein J, Lavery R. Energetic coupling between DNA bending and base pair opening. Proc Natl Acad Sci USA 1988;85:7231–7235.
27. Mehler EL, Eichele G. Electrostatic effects in water accessible regions of proteins. Biochemistry 1984;23:3887–3892.
28. Hassan SA, Guarnieri F, Mehler EL. A general treatment of solvent effects based on screened coulomb potentials. J Phys Chem B 2000;104:6478–6489.
29. Hassan SA, Mehler EL. A critical analysis of continuum electrostatics: the screened coulomb potential-implicit solvent model and the study of the alanine dipeptide and discrimination of misfolded structure of proteins. Proteins 2002;47:45–61.
30. Allen MP, Tildesley DJ. Computer simulation of liquids. Oxford: Oxford University Press; 1987
31. Pastor RW, Karplus M. Parametrization of the friction constant for stochastic simulations of polymers. J Phys Chem 1988;92:2636–2641.
32. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A & Haak JR. Molecular dynamics with coupling to an external bath. J Chem Phys 1984;81:3684–3690.
33. Press WH, Teukolsky SA, Vetterling WT, Flannery BP. Numerical recipes. Cambridge: Cambridge University Press; 1992.
34. Kabsch W & Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 1983;22:2577–2637.
35. Gasteiger J, Marsili M. Iterative partial equalization of orbital electronegativity: a rapid access to atomic charges. Tetrahedron Lett 1980;36:3219–3228.