Atomically detailed folding simulation of the B domain of staphylococcal protein A from random structures

Jorge A. Vila*†, Daniel R. Ripoll‡, and Harold A. Scheraga*§

*Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, NY 14853-1301; †Facultad de Ciencias Físico Matemáticas y Naturales, Instituto de Matemática Aplicada San Luis, Consejo Nacional de Investigaciones Científicas y Técnicas de Argentina, Universidad Nacional de San Luis, Ejército de Los Andes 950-5700 San Luis, Argentina; and ‡Computational Biology Service Unit, Cornell Theory Center, Cornell University, Ithaca, NY 14853-3801

Contributed by Harold A. Scheraga, October 6, 2003

The conformational space of the 10–55 fragment of the B-domain of staphylococcal protein A has been investigated by using the electrostatically driven Monte Carlo (EDMC) method. The ECEPP/3 (empirical conformational energy program for peptides) force-field plus two different continuum solvation models, namely SRFOPT (Solvent Radii Fixed with atomic solvation parameters OPTimized) and OONS (Ooi, Oobatake, Némethy, and Scheraga solvation model), were used to describe the conformational energy of the chain. After an exhaustive search, starting from two different random conformations, three of four runs led to native-like conformations. Boltzmann-averaged root-mean-square deviations (RMSD) for all of the backbone heavy atoms with respect to the native structure of 3.35 Å and 4.54 Å were obtained with SRFOPT and OONS, respectively. These results show that the protein-folding problem can be solved at the atomic detail level by an *ab initio* procedure, starting from random conformations, with no knowledge except the amino acid sequence. To our knowledge, the results reported here correspond to the largest protein ever folded from a random conformation by an initial-value formulation with a full atomic potential, without resort to knowledge-based information.

For many years, methods have been developed to compute the 3D structures of polypeptides, based on empirical atomic-based potential energy functions and global optimization of such functions. Because of limitations in computer power, these methods have been confined to small molecules such as the pentapeptide enkephalin (1–15), the decapeptide gramicidin S (16–21), and linear fibrous proteins such as collagen-like repeating polytripeptides (22–24). Inclusion of explicit or implicit hydration in the potential function only exacerbated the global optimization problem. However, with the recent availability of cost-effective alternatives to large supercomputers, such as Beowulf class cluster computers (25), it is now possible to extend the application of such *ab initio* physics-based methods to larger molecules.

In this article, we report the results of the global optimization of the all-atom force field ECEPP/3 (empirical conformational energy program for peptides) (26–29) plus two implicit hydration models [SRFOPT (Solvent Radii Fixed with atomic solvation parameters OPTimized; ref. 30) and OONS (Ooi, Oobatake, Némethy, and Scheraga solvation model; ref. 31)], using the electrostatically driven Monte Carlo (EDMC) method (32, 33) to explore the conformational space of the 10–55 fragment of the B-domain of the staphylococcal protein A molecule, efficiently. The structure of this fragment of the B-domain of the protein A molecule is known from x-ray (34) and NMR (35) investigations, and from minimalist and all-atom simulations (36–46). However, such initial-value-formulated simulations (except for ref. 46, which is a boundary-value formulation) were not started from a random conformation. Therefore, in this work, we attempted to provide an extensive exploration of the conformational space by starting from two different randomly chosen conformations, using a Beowulf class cluster.

Methods

Evaluation of the Conformational Energy. The conformations were generated with the EDMC method (32, 33). An all-atom representation of the chain was used with the ECEPP/3 force field (26–29). Two alternative forms of the potential energy function were used to evaluate the total energy, $E(\mathbf{r}_p)$, as a function of the coordinates $\mathbf{r}_p$, namely,

$$E(\mathbf{r}_p) = E_{int}(\mathbf{r}_p) + F_{sas}(\mathbf{r}_p), \quad [1]$$

where $E_{int}(\mathbf{r}_p)$ is the internal conformational energy of the molecule in the absence of solvent, assumed to correspond to the ECEPP/3 energy of the neutral molecule and $F_{sas}(\mathbf{r}_p)$ represents the solvation free energy as defined by Vila *et al.* (30). Two different solvation models were used during the simulations, namely, SRFOPT (30) and OONS (31).

The Starting Point: Generation of the Random Conformation. The N and C termini of the 10–55 fragment of the B-domain of the staphylococcal protein A molecule were blocked by amino-$COCH_3$ and carboxyl-NH_2 groups, respectively. The simulations were started from two different initial random conformations, i.e., namely rnd_1 and rnd_2. For each of these random conformations, *all* backbone and side-chain dihedral angles were chosen randomly between 180.0° and −180.0°, with the exception of the dihedral angles ω of the peptide group, which were always chosen in the *trans* (180.0°) conformation. All backbone and side-chain dihedral angles (including ω's) were allowed to vary freely during the simulations, with the exception of proline residues. In the ECEPP/3 fixed geometry approximation, both up (U) and down (D) puckering conformations of the pyrrolidine ring (which pertain to the following: $\phi = -53.0°$ and $\chi^1 = -28.1°$ and $\phi = -68.8°$ and $\chi^1 = 27.4°$ positions, respectively, of the C^γ atom of the proline residue) are considered. At the beginning of a simulation, the puckering of the proline residues is chosen randomly. However, up and down puckering and *cis* ↔ *trans* isomerization of the peptide group preceding proline residues were allowed to permute during the course of the simulations. During the EDMC search, if a proline residue is selected for a change, an attempt is made to change the puckering state with a probability of 1/2. In addition, the

Abbreviations: ECEPP/3, empirical conformational energy program for peptides; EDMC, electrostatically driven Monte Carlo; RMSD, rms deviation; SRFOPT, solvent radii fixed with atomic solvation parameters OPTimized; OONS, Ooi, Oobatake, Némethy, and Scheraga solvation model.

§To whom correspondence should be addressed. E-mail: has5@cornell.edu.

www.pnas.org/cgi/doi/10.1073/pnas.2436463100

Table 1. Summary of the EDMC runs*

Run	Solvent model	No. of energy-minimized conformations†	No. of accepted conformations‡	Lowest energy, kcal/mol	Total computational time, hr§
1¶	SRFOPT	2,168,453	14,654	−715.5	262
2‖	SRFOPT	2,199,341	14,539	**−720.7**	191
3**	OONS	815,002	6,546	−651.4	80
4††	OONS	431,214	3,045	*−652.9*	50
5‡‡	SRFOPT	247,471	1,840	*−702.5*	47
6§§	OONS	402,627	3,075	**−654.3**	47

*Using an average of 60 processors for each run.
†Number of generated conformations, using the procedure described in *Methods.*
‡According to the Metropolis criterion.
§Differences in computational time reflect different starting conformations and different judgments as to when the runs converged.
¶Run 1 was started from random conformation rnd_1.
‖Run 2 was started from random conformation rnd_2. The lowest-energy identified with this solvation model is indicated in boldface type and belongs to a native-like structure.
**Run 3 was started from random conformation rnd_1.
††Run 4 was started from random conformation rnd_2, which led to a mirror image of the native fold. The energy of the mirror-image conformation is indicated in italics.
‡‡Crosscheck test. Run 5 was started from the lowest energy conformation described in ** and run subsequently with the SRFOPT solvation model. Energy of the mirror-image conformation is indicated in italics.
§§Crosscheck test. Run 6 was started from the lowest-energy conformation described in ¶ and run subsequently with the OONS solvation model. The lowest-energy identified with this solvation model is indicated in boldface type and belongs to a native-like structure.

probability of changing the dihedral angles ω preceding proline from *trans* to *cis* is 1/3, and the values of ω were allowed to vary around the *trans* and *cis* conformations.

The Conformational Search. For each rnd_1 and rnd_2 initial conformation, an ensemble of conformations was generated by using the EDMC method, as shown in Table 1 for runs 1–4. During each of these runs, the energies of the initially generated and subsequent conformations were minimized by using the secant unconstrained minimization solver (SUMSL) algorithm (47) in combination with ECEPP/3 plus a surface accesible solvation model, either SRFOPT or OONS. Only a small set of low-energy conformations was retained. This set corresponded to the accepted conformations from the Monte Carlo path followed by the EDMC method in each of these runs. It should be noted that the acceptance rate (see Table 1) is low for all of the runs in our simulations. However, as was noted previously (48, 49), this acceptance rate is characteristic of the EDMC procedure.

Evaluation of the Results: Computation of the Root-Mean-Square Deviation (RMSD). We chose to compare the results of our simulations with the minimized-averaged structure of the 10–55 fragment of the B-domain of the staphylococcal protein A molecule free in solution determined from NMR at 30°C by Gouda *et al*. (35), and not with the three-dimensional structure of the 10–55 fragment of the B-domain of the staphylococcal protein A molecule bound to the Fc fragment of human polyclonal IgG, determined by x-ray crystallographic analysis by Deisenhofer (34). The reason for this decision was that the B-domain of the staphylococcal protein A molecule forms contacts with the Fc fragment in the crystal, and this could be the main source of the structural differences observed between the NMR and x-ray structures, i.e., the fragment¶ Ser-42–Ala-55 constitutes helix III of the NMR structure whereas, in the crystallographic structure, there is no structural information available for the segment Ala-49–Lys-59, and the portion of the polypeptide chain from Ser-42–Glu-48 is in an extended conformation.

¶Residue numbers correspond to those used for the NMR form in the Protein Data Bank: file 1BDD.

Because all of the NMR constraints used to determine the most probable conformation in solution represent population-weighted-averaged measurements, we computed the Boltzmann-averaged values for the RMSD from the minimized-averaged structure obtained from the NMR experiment (35), using the *separate* ensembles of conformations generated from each of two different starting random conformations.

Results

Figs. 1 and 2 show a scatter plot of the C^{α} RMSDs from the native structure for *all* the accepted conformations listed in Table 1 for the runs with the SRFOPT and OONS surface area model, respectively. Table 2 shows the Boltzmann-averaged RMSDs from the native fold obtained with these solvation models for *all* the accepted conformations obtained for each of the simulations

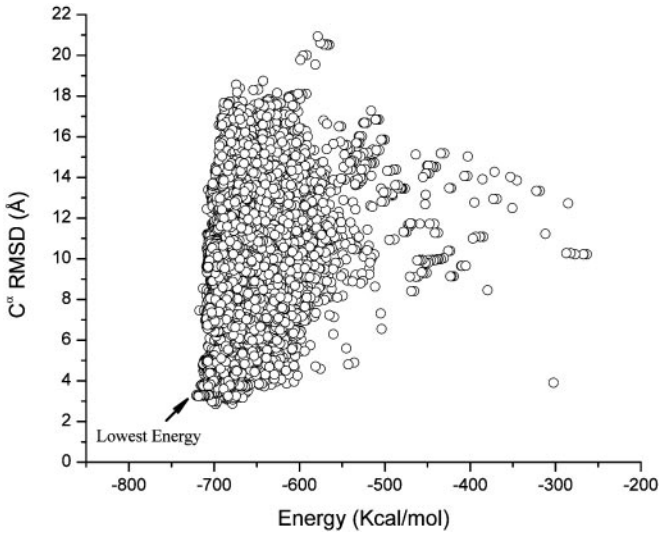

Fig. 1. C^{α} RMSD vs. the total energy, computed by Eq. **1** using the SRFOPT solvation model, for each of the accepted conformations obtained during runs 1 and 2. The scatter plot does not include those conformations obtained after the crosscheck (run 5). The lowest-energy identified in the figure belongs to the lowest-energy structure (−720.7 kcal/mol) identified in *all* the runs carried out with the SRFOPT model, namely, runs 1, 2, and 5.

BIOPHYSICS

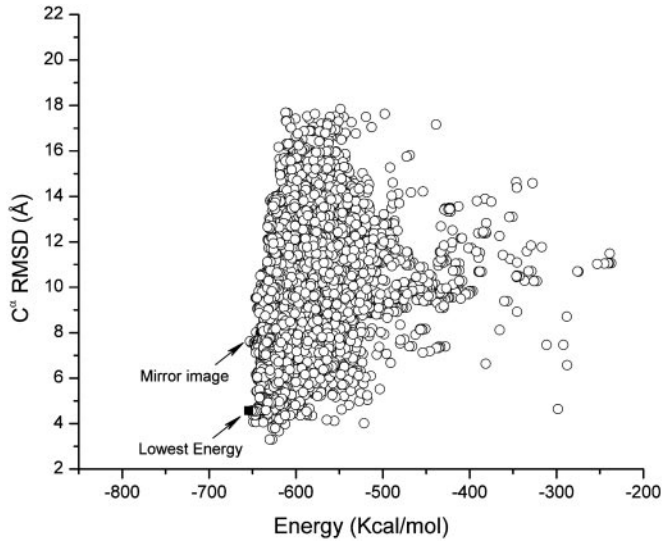

Fig. 2. C^{α} RMSD vs. the total energy, computed by Eq. **1** using the OONS solvation model, for each of the accepted conformations obtained during runs 3 and 4. The scatter plot does not include those conformations obtained after the crosscheck (run 6). The mirror image (−652.9 kcal/mol) of the native fold identified in the figure belongs to the lowest-energy found in runs 3 and 4. However, this conformation is higher in energy than those obtained after the crosschecking test in run 6 (−654.3 kcal/mol); the conformation in run 6 belongs to the native-like fold and is indicated by a filled square in the figure and represents the lowest-energy conformation identified in all the runs carried out with the OONS solvation model, namely, runs 3, 4, and 6.

starting from rnd_1 and rnd_2. The plots in Figs. 1 and 2 pertain to the C^{α} (rather than the backbone heavy atoms or all of the heavy atoms) RMSDs from the native structure because there are very good correlations between these quantities and the C^{α} RMSDs, with a correlation coefficient of $R = 0.997$ and 0.980, respectively. In other words, plots of the RMSDs of the backbone heavy atoms or all of the heavy atoms, instead of the C^{α} rms deviations from the native structure, do not add any new information.

The NMR-determined structure of the 10–55 fragment of the B-domain of the staphylococcal protein A molecule is composed of a bundle of three α-helices as shown in Fig. 3*b*, i.e., helix I (Gln-10–His-19), helix II (Glu-25–Asp-37), and helix III (Ser-42–Ala-55). The lowest energy conformation found with SRFOPT (−720.7 kcal/mol) exhibits a similar distribution of helices, i.e., helix I (Gln-11–His-19), helix II (Arg-28–Asp-37), and helix III (Pro-39–Asp-54). Helices II and III are antiparallel

Table 2. Boltzmann-averaged RMSD* between native and generated conformations

Solvent model	C^{α} atoms, Å	Backbone heavy atoms, Å	All heavy atoms, Å
SRFOPT†	3.96 (2.85)	4.05 (2.92)	5.29 (4.21)
SRFOPT‡	3.30 (3.24)	3.35 (3.30)	4.69 (4.62)
OONS§	4.09 (3.31)	4.54 (3.42)	5.46 (4.61)
OONS¶	7.59 (4.35)	7.59 (4.42)	8.91 (5.42)

*Computed at 30°C by using *separate* ensembles of conformations generated in runs 1 to 4, listed in Table I. Values within parentheses belong to the lowest RMSD computed for a single conformation, i.e., they do not represent Boltzmann-averaged values.
†Results from run 1, Table 1.
‡Results from run 2, Table 1.
§Results from run 3, Table 1.
¶Results from run 4, Table 1, which led to a mirror image of the native fold.

to each other in both the NMR-derived native fold and in the lowest-energy conformation found with either SRFOPT or OONS, as shown in Fig. 3 *c* and *d*.

Independent of the starting conformation, the SRFOPT solvation model always led to a native-like structure. Both of the lowest-energy conformations identified in these runs (−715.5 and −720.7 kcal/mol, respectively) have native-like folds. Analysis of the runs from Table 1 shows that the potential function with the OONS solvation parameters led to two broad basins with similar energies. One of the basins contains the native-like structure (−651.4 kcal/mol) and the second one contains conformations that represent *mirror* images of the native-like structure (RMSD vs. Energy is indicated in Fig. 2).

To investigate the depth of the basins containing the native- and mirror-like conformations, we carried out additional crosschecking runs. The lowest-energy conformation obtained with SRFOPT in run 2 (−720.7 kcal/mol) was used as the starting conformation with the OONS solvation model and, vice versa, the lowest-energy conformation, i.e., the mirror image, obtained with OONS in run 4 (−652.9 kcal/mol) was used as the starting conformation with the SRFOPT solvation model. The resulting total energy shown in Table 1 for the mirror image (in italics) indicates that, for both solvation models, the mirror image is *higher* in energy than those belonging to the native-like conformation, after the crosschecking. It should be noted that no constraints were used to restrict the search during these runs.

It was found that the potential energy with OONS seems to distinguish the native-like structure (−654.3 kcal/mol) from the mirror image (−652.9 kcal/mol); however, the energy difference between these two basins is quite small. The potential energy with SRFOPT, on the other hand, shows a bigger difference between both structures, namely, −720.7 kcal/mol for the native-like vs. −702.5 kcal/mol for the mirror image structure, indicating that this solvation model can detect the correct fold. A summary of these crosschecking runs is shown in Table 1, runs 5 and 6.

The scatter plots shown in Figs. 1 and 2 indicate that the differences in energy among the few conformations close to the lowest one are small whereas they display a broad dispersion in RMSD. This result represents one of the main difficulties in identifying the native-like structure. On the other hand, this result means that the total energy, as a scoring function, must be extremely precise to distinguish the correct folded structure from wrong ones. This requirement constitutes a challenge for improving existing force fields or for developing new potential functions. In addition, the small energy gap between basins containing quite different folds represents a great challenge for search methods. Regarding this observation, it can be seen from Table 2 that the EDMC method did find a conformation that is quite close to the native structure (with an RMSD for the C^{α} atoms of 2.85 Å). However, this conformation with the lowest RMSD (−698.3 kcal/mol) is higher in energy than the lowest energy minimum (−720.7 kcal/mol) found for this solvation model by >20 kcal/mol. Fig. 4 shows the superposition of this structure with the native-NMR fold. It should be pointed out, for comparative purposes only, that the C^{α} RMSD of the fragment containing helix I, helix II, and part of helix III, i.e., residues‖ 128–162 of the x-ray (35) and equivalent residues 10–44 of the NMR (36) structures, is 2.0 Å.

Discussion and Conclusions

There is an ongoing interest about how to represent the unfolded state of a protein. According to some authors (50), the denatured

‖Residue numbers correspond to those used for the crystalline form in the Protein Data Bank: file 1FC2.

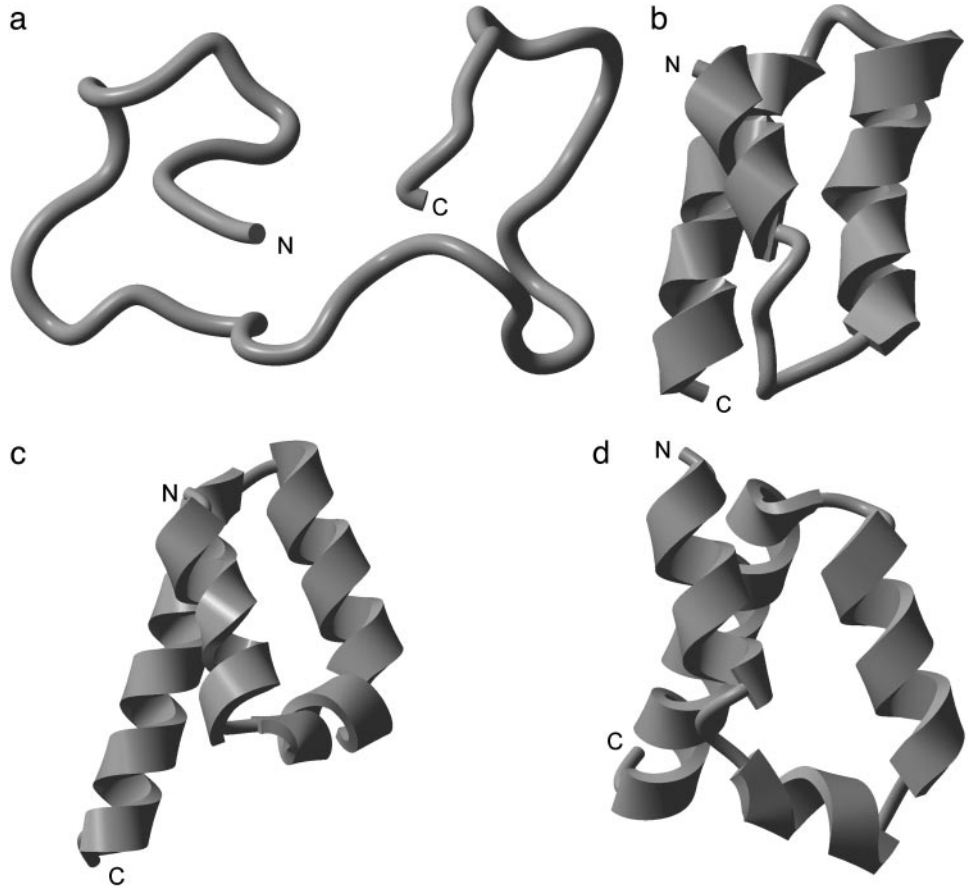

Fig. 3. (*a*) Ribbon diagram of the starting conformation rnd_1. (*b*) Ribbon diagram of the native fold obtained by NMR. (*c*) Ribbon diagram of the lowest-energy (−720.7 kcal/mol) structure identified in this work by using the SRFOPT solvation model. (*d*) Ribbon diagram of the lowest-energy (−654.3 kcal/mol) structure identified in this work by using the OONS solvation model.

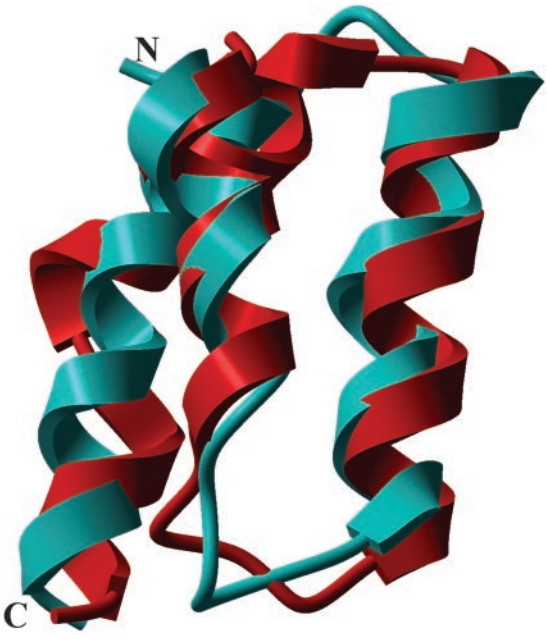

Fig. 4. Ribbon diagram of the superposition of the native fold (cyan) and the conformation (red) with the lowest C^{α} RMSD (2.85 Å) from the native fold, obtained with the SRFOPT solvation model. The energy of this conformation is >20 kcal/mol above that of the lowest-energy conformation identified with this solvation model in runs 1, 2, and 5.

states are quite sensitive to their environment and hence they can span a continuum of conformational states, ranging from structured intermediates to statistical coils. In other words, it seems that the residual structure depends on the sequence and the environment. Correct characterization of the unfolded states represents a more complex problem than the protein-folding problem by itself because the unfolded state may not correspond to a well-defined structure such as that characterizing the native fold of a protein. For this reason, and also to avoid any *bias* in the conformational search, we started from random initial conformations with no knowledge other than the amino acid sequence, i.e., an *ab initio* search. Our results show that folding of a small protein, such as the 10–55 fragment of the B-domain of the staphylococcal protein A molecule can be accomplished with an all-atom potential with a reasonable degree of accuracy, as shown in Table 2. It is worth noting that other contributions to the total energy, such as coupling of the conformation with the state of ionization of the individual amino acids (proton-binding equilibrium) (51) or an estimation of the conformational entropy (52), were not taken into account in these simulations. Conceivably, incorporation of these components as well as a new parameterization of the existing all-atom force-field (53) should improve the accuracy of the predictions.

We wrote in 1994 (54), "With the recent development of various efficient approaches to overcome the multiple-minima problem, the problem has been in some sense solved for small oligopeptides, as well as for regular-repeating structures and assemblies of fibrous proteins. It can reasonably be expected that the current extension of these to globular proteins will result in

efficient searches of their conformational space in the near future. Advances in computer hardware and software, especially the wider use of parallelism, will speed up computations, making practical the application to larger molecules. The point may have been reached where wider applicability of these techniques is becoming limited by the accuracy of the potential energy functions used to describe the energetic of polypeptide and protein structure." Results of this work, in some sense, support this assertion. The 46-residue fragment of protein A is larger than the 36-residue α-helical protein from the villin headpiece, for which all-atom simulations starting from an extended structure were previously carried out (55, 56). It is worth noting that simulations on the villin headpiece were carried out with explicit solvent (55), which increases the computing time considerably compared with the time required for the implicit solvent models used in our simulations. However, despite the large progress in computer technology, we still cannot extend the all-atom approach to the prediction of the folding of globular proteins containing 100–200 residues. For this reason, the use of simpler models based on physical grounds, such as the hierarchical approach starting with the united-residue (UNRES) force field and finishing with an all-atom search (57), is necessary. Also, procedures that combine knowledge-based information with energy minimization (58–60) have achieved some measure of success in recent CASP (*c*ritical *a*ssessment of techniques for protein *s*tructure *p*rediction) blind tests. Hopefully, further improvements will extend the applicability of all-atom *ab initio* procedures.

We thank Drs. A. Liwo and Y. Arnautova for valuable discussion. This research was supported by National Institutes of Health Grants GM-14312 and TW00857 and National Science Foundation Grant MCB00-03722. Support was also received from the National Foundation for Cancer Research, the National Research Council of Argentina [Consejo Nacional de Investigaciones Científicas y Técnicas de Argentina (CONICET)], and Project P-328402 of the Universidad Nacional de San Luis, Argentina. Part of this research was carried out by using (*i*) the resources of the Cornell Theory Center, which receives funding from Cornell University, New York State, federal agencies, foundations, and corporate partners and (*ii*) the National Partnership for Advanced Computational Infrastructure at the Pittsburgh Supercomputing Center, which is supported in part by National Science Foundation Grant MCA99S007P.

1. Isogai, Y., Némethy, G. & Scheraga, H. A. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 414–418.
2. Paine, G. P. & Scheraga, H. A. (1985) *Biopolymers* **24,** 1391–1436.
3. Vasquez, M. & Scheraga, H. A. (1985) *Biopolymers* **24,** 1437–1447.
4. Paine, G. H. & Scheraga, H. A. (1987) *Biopolymers* **26,** 1125–1162.
5. Li, Z. & Scheraga, H. A. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 6611–6615.
6. Li, Z. & Scheraga, H. A. (1988) *J. Mol. Struct. (Theochem)* **179,** 333–352.
7. Ripoll, D. R. & Scheraga, H. A. (1989) *J. Protein Chem.* **8,** 263–287.
8. Nayeem, A., Vila, J. & Scheraga, H. A. (1991) *J. Comput. Chem.* **12,** 594–605.
9. Abagyan, R. & Argos, P. (1992) *J. Mol. Biol.* **225,** 519–532.
10. Androulakis, I. P., Maranas, C. D. & Floudas, C. A. (1997) *J. Glob. Optimization* **11,** 1–34.
11. van der Spoel, D. & Berendsen, H. J. C. (1997) *Biophys. J.* **72,** 2032–2041.
12. Hansmann, U. H. E. & Okamoto, Y. (1999) *Braz. J. Phys.* **29,** 187–198.
13. Klepeis, J. L. & Floudas, C. A. (1999) *J. Chem. Phys.* **110,** 7491–7512.
14. Kinoshita, M., Okamoto, Y. & Hirata, F. (2000) *J. Am. Chem. Soc.* **122,** 2773–2779.
15. Sanbonmatsu, K. Y. & Garcia, A. E. (2002) *Proteins Struct. Funct. Genet.* **46,** 225–234.
16. Vanderkooi, G., Leach, S. J., Némethy, G., Scott, R. A. & Scheraga, H. A. (1966) *Biochemistry* **5,** 2991–2997.
17. Scott, R. A., Vanderkooi, G., Tuttle, R. W., Shames, P. M. & Scheraga, H. A. (1967) *Proc. Natl. Acad. Sci. USA* **58,** 2204–2211.
18. De Santis, P. & Liquori, A. M. (1971) *Biopolymers* **10,** 699–710.
19. Dygert, M., Go, N. & Scheraga, H. A. (1975) *Macromolecules* **8,** 750–761.
20. Naik, V. M., Krimm, S., Denton, J. B., Némethy, G. & Scheraga, H. A. (1984) *Int. J. Pept. Protein Res.* **24,** 613–626.
21. Mihailescu, D. & Smith, J. C. (2000) *Biophys. J.* **79,** 1718–1730.
22. Miller, M. H. & Scheraga, H. A. (1976) *Polymer Symposia* **54,** 171–200.
23. Miller, M. H., Némethy, G. & Scheraga H. A. (1980) *Macromolecules* **13,** 470–478.
24. Vitagliano, L., Némethy, G., Zagari, A. & Scheraga, H. A. (1995) *J. Mol. Biol.* **247,** 69–80.
25. Sterling, T. & Savarese, D. (1999) *Lecture Notes Comput. Sci.* **1685,** 78–88.
26. Momany, F. A., McGuire, R. F., Burgess, A. W. & Scheraga H. A. (1975) *J. Phys. Chem.* **79,** 2361–2381.
27. Némethy, G., Pottle, M. S. & Scheraga H. A. (1983) *J. Phys. Chem.* **87,** 1883–1887.
28. Sippl, M. J., Némethy, G. & Scheraga, H. A. (1984) *J. Phys. Chem.* **88,** 6231–6233.
29. Némethy, G., Gibson, K. D., Palmer, K. A., Yoon C. N., Paterlini G., Zagari, A., Rumsey, S. & Scheraga H. A. (1992) *J. Phys. Chem.* **96,** 6472–6484.
30. Vila, J., Williams, R. L., Vásquez, M. & Scheraga, H. A. (1991) *Proteins Struct. Funct. Genet.* **10,** 199–218.
31. Ooi, T., Oobatake, M., Némethy, G. & Scheraga, H. A. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 3086–3090.
32. Ripoll, D. R. & Scheraga, H. A. (1988) *Biopolymers* **27,** 1283–1303.
33. Ripoll, D. R., Liwo, A. & Czaplewski, C. (1999) *TASK Q.* **3,** 313–331.
34. Deisenhofer, J. (1981) *Biochemistry* **20,** 2361–2370.
35. Gouda, H., Torigoe, H., Saito, A., Sato, M., Arata, Y. & Shimada, I. (1992) *Biochemistry* **31,** 9665–9672.
36. Kolinski, A. & Skolnick, J. (1994) *Proteins Struct. Funct. Genet.* **18,** 353–366.
37. Boczko, E. M. & Brooks, C. L., III (1995) *Science* **269,** 393–396.
38. Onuchic, J. N., Wolynes, P. G., Luthey-Schulten, Z. & Socci, N. D. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 3626–3630.
39. Guo, Z., Brooks, C. L., III & Boczko, E. M. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 10161–10166.
40. Shea, J. E., Onuchic, J. N. & Brooks, C. L., III (1999) *Proc. Natl. Acad. Sci, USA* **96,** 12512–12517.
41. Zhou, Y. & Karplus, M. (1999) *J. Mol. Biol.* **293,** 917–951.
42. Takada, S., Luthey-Schulten, Z. & Wolynes, P. G. (1999) *J. Chem. Phys.* **110,** 11616–11629.
43. Lee, J., Liwo, A. & Scheraga, H. A. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 2025–2030.
44. Alonso, D. O. V. & Daggett, V. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 133–138.
45. Favrin, G., Irbäck, A. & Wallin, S. (2002) *Proteins Struct. Funct. Genet.* **47,** 99–105.
46. Ghosh, A., Elber, R. & Scheraga, H. A. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 10394–10398.
47. Gay, D. M., (1983) *ACM Trans. Math. Software* **9,** 503–524.
48. Vila, J., Williams, R. L., Grant, J. A., Wojcik, J. & Scheraga, H. A. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 7821–7825.
49. Vila, J. A., Ripoll, D. R. & Scheraga, H. A. (2001) *Biopolymers* **58,** 235–246.
50. Daggett, V. & Fersht, A. (2003) *Nat. Rev. Mol. Cell Biol.* **4,** 497–502.
51. Ripoll, D. R., Vorobjev, Y. N., Liwo, A., Vila, J. A. & Scheraga, H. A. (1996) *J. Mol. Biol.* **264,** 770–783.
52. Gō, N. & Scheraga, H. A. (1969) *J. Chem. Phys.* **51,** 4751–4767.
53. Arnautova, Y. A., Jagielska, A. Pillardy, J. & Scheraga, H. A. (2003) *J. Phys. Chem. B* **107,** 7143–7154.
54. Vásquez, M., Némethy, G. & Scheraga, H. A. (1994) *Chem. Rev.* **94,** 2183–2239.
55. Duan, Y. & Kollman, P. A. (1998) *Science* **282,** 740–744.
56. Zagrovic, B., Snow, C. D., Shirts, M. R. & Pande, V. S. (2002) *J. Mol. Biol.* **323,** 927–937.
57. Liwo, A., Lee, J., Ripoll, D. R., Pillardy, J. & Scheraga, H. A. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 5482–5485.
58. Alm, E. & Baker, D. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 11305–11310.
59. Eyrich, V. A., Standley, D. M. & Friesner, R. A. (2002) *Adv. Chem. Phys.* **120,** 223–264.
60. Skolnick, J. & Kolinski A. (2002) *Adv. Chem. Phys.* **120,** 131–192.