Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 709–713, January 1995
Biophysics

Discriminating compact nonnative structures from the native structure of globular proteins

(protein folding/tertiary structure/conformational sampling/molecular dynamics/atomic solvation)

YUHONG WANG*†, HUI ZHANG†‡, WEI LI*, AND ROBERT A. SCOTT‡

*Department of Molecular Biology, Jilin University, Changchun 130023, People's Republic of China; and ‡Center for Metalloenzyme Studies, University of Georgia, Athens, GA 30602-2556

Communicated by Norman L. Allinger, University of Georgia, Athens, GA, October 25, 1994 (received for review May 10, 1994)

ABSTRACT **Prediction of the native tertiary structure of a globular protein from the primary sequence will require a potential energy model that can discriminate all nonnative structures from the native structure(s). A successful model must distinguish not only alternate structures that are very nonnative but also alternate structures that are compact and near-native. We describe here a method, based on molecular dynamics simulation, that allows generation of hundreds of compact alternate structures that are arbitrarily close to the native structure. In this way, a significant amount of conformational space in the neighborhood of the native structure can be sampled and these alternate structures can be used as a stringent test of protein folding models. We have used two sets of these alternate structures generated for six crystallographically characterized small globular proteins (1200 alternate structures in all) to test eight empirical energy models for their ability to discriminate alternate from native structures. Seven of the models fail to correctly identify at least some of the alternate structures as nonnative. An atomic solvation model is presented that succeeds in discriminating all 1200 alternate structures from native.**

If it is assumed that the native structure of a globular protein is controlled by thermodynamics, then the *ab initio* prediction of native tertiary structure from protein amino acid sequence will require (*i*) the identification of the free energy hypersurface on which the native structure is at the global minimum and (*ii*) designing efficient algorithms for exploring enough conformational space to find this global minimum. Herein we describe a method for testing potential energy functions for their ability to discriminate the native structure (the global minimum) from alternate nonnative structures (local minima). We have developed an empirical potential energy function based on atomic solvation that is here compared to other existing functions and found to be highly discriminating against such alternate structures.

Traditional molecular mechanics force fields such as AMBER (1, 2) and CHARMM (3) are unable to discriminate between native and alternate (nonnative) structures (4, 5). For this reason a number of empirical potential energy functions have been developed and employed in lattice-model simulations of protein folding (6–12), identification of distantly related proteins (13–19), and recognition of errors in three-dimensional protein structures (20). These empirical models can be classified into the following four groups: (*i*) contact models (10, 18, 21–24) in which amino acid residues are simplified to "super atoms" and inter-super-atom contact energies drive folding to the native structure; (*ii*) mean force models (14, 19, 25–28) in which thousands of interaction energy parameters are derived for each pair of residues at different spatial and sequential separations; (*iii*) local models (13, 24, 29–36) in which preferences of individual residues for particular dihedral angles, solvent exposure, or secondary structure direct folding; and (*iv*) models that combine these three strategies (16, 17, 37, 38).

Testing of such empirical potential energy models requires a collection of alternate (nonnative) protein structures; a good model should yield a higher potential energy for all alternate structures compared to native. A stringent test would use alternate structures that are themselves compact, globular, and not very different from native; a model that can discriminate these closely similar structures would still help direct native folding when the tertiary structure gets close to native. A common method for generating alternate structures is based on the sequence-recognizes-structure (SRS) protocol developed by Hendlich *et al.* (25). The sequence of interest is modeled as a sequence fragment of all larger protein structures and is threaded through each larger protein sequence advancing one residue at a time, generating a large number of alternate structures (14, 18, 19, 23, 24, 27, 28, 37). Disadvantages to this method are that the structures generated are very nonnative (they are usually not compact) and that side chains are often disregarded. The tests that we perform herein make use of a collection of hundreds of alternate structures generated by starting with a native structure and using molecular dynamics (MD) to sample the nearby conformational space and molecular mechanics energy minimization to place the alternate structures into local minima. These stringent tests allow successful empirical potential energy models to be identified.

MATERIALS AND METHODS

Test Protein Structures. Six proteins of moderate size were chosen from the Protein Data Bank for generating test structures; they were chosen to be high resolution (≤2.0Å), to have all nonhydrogen atomic coordinates identified, and to lack prosthetic groups. The proteins used were as follows: crambin, 1crn (39); ovomucoid protein, 2ovo (40); trypsin inhibitor, 4pti (41); scorpion neurotoxin, 1sn3 (42); amylase inhibitor, 1hoe (43); and ubiquitin, 1ubq (44). The crystal structure coordinates were subjected to 1000 cycles of AMBER 4.0 (1) molecular mechanics energy minimization using the united-atom model to generate the "native" structures [which differed from crystal structures by rms deviations (rmsds) of 0.2–0.5 Å; all rmsds are calculated according to Kabsch (45, 46) and are based on C_α coordinates]. All calculations were performed on an IBM RISCStation 6000 cluster supported by University Computing and Networking Services of the University of Georgia.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: SRS, sequence recognizes structure; rmsd, rms deviation; AM, AMBER model; MC, Maiorov and Crippen model; MJ, Miyazawa and Jernigan model; BE, Bowie *et al.* model; HS, Hendlich *et al.* model; GS, Godzik and Skolnick model; LS, Holm and Sander model; WZS, Wang, Zhang, and Scott model; MD, molecular dynamics model.

†To whom reprint requests should be addressed.

Generation of Alternate Structures. All simulations described here used the AMBER 4.0 united-atom force field for *in vacuo* simulations with a constant dielectric ($\varepsilon = 4$), a nonbonded cutoff of 9.5 Å, and a time step of 1 fs. The SHAKE algorithm maintained all bond lengths at their equilibrium value (±0.0004 Å). For each test protein, the native (energy-minimized) structure was heated from 0 K to the desired temperature (300 or 500 K) in increments of 30 K applied every 200 fs. The resulting structure was equilibrated through an additional 5 ps by rerandomizing each atom's velocity according to a Boltzmann distribution. During the ensuing data collection, structures were saved every 4 ps for a total of 400 ps (100 structures) at each temperature, generating a total of 1200 alternate structures for the six proteins.

Models Tested. Eight potential energy models were tested for their ability to discriminate alternate from native structures as follows: AM, the AMBER 4.0 molecular mechanics force field (1, 2) was used to calculate the potential energies of each structure; MC, the contact potential energy function of Maiorov and Crippen (23) was used with parameter values from their table 5; MJ, the contact energy function of Miyazawa and Jernigan (22) was used with parameter values from their table V; BE, the three-dimensional profile method of Bowie *et al.* (13) was performed using the program kindly provided by the authors; HS, the potential function of mean force (25, 26) was used in its most recent form and the results in the form of *z* scores (27) were kindly provided by M. J. Sippl; GS, the topology fingerprint approach of Godzik and Skolnick (16) was used in its most recent form and the results were kindly provided by A. Godzik; LS, the atomic solvation preference model of Holm and Sander (47) was used with parameter values from their table 1.

Existing atomic solvation parameters (30, 31, 48) poorly discriminate alternate from native structures, so we developed a modified atomic solvation model, referred to as WZS (Wang, Zhang, and Scott model). In this model, protein atoms are classified into 14 groups (see below) and each group is assigned a separate solvation parameter (σ_i). The solvent-accessible surface area (A_i) is calculated according to a fast algorithm (49) by using a 1.4-Å water radius and the solvation energy is calculated as $\Sigma_i \sigma_i A_i$. The native structure and alternate structures at 500 K of two arbitrarily chosen proteins were used to train the 14 atomic solvation parameters. Three sets of solvation parameters were generated using different pairs of proteins in the training set. WZS1 used 1sn3 and 1ubq, WZS2 used 1hoe and 1ubq, and WZS3 used 1sn3 and 1hoe. In the training process, randomly initialized parameters were adjusted, according to the back-propagation algorithm (50), to maximize the solvation energy differences between native and alternate structures. A detailed description of this solvation model will be presented elsewhere (Y.W., H.Z., and R.A.S., unpublished data).

To compare models, each of which generates energies on different absolute scales, we calculated, for each alternate structure, a scaled energy difference from native, Δ_i,

$$\Delta_i = \frac{E^{\text{native}} - E_i^{\text{alternate}}}{\max(|E^{\text{native}}|, |E_i^{\text{alternate}}|)}, \quad [1]$$

where E^{native} and $E^{\text{alternate}}$ are the empirical energies [or *z* scores for the HS model or negative quality scores (*QS*) for the BE model] of the native and alternate structures, respectively. The means and estimated standard deviations of Δ_i over all structures at 300 K or 500 K were also calculated to give a gross comparison of the models.

RESULTS

Table 1 summarizes characteristics of the six test proteins and the alternate structures generated from them. The average radii of gyration of the alternate structures [R_g(alt)] range from 93 to 107% of that of the native structure [R_g(nat)] and average 95%. These structures are as compact as the native structure in each case. The alternate structures generated at 300 K have mean rmsds from native ranging from 2.8 to 4.2 Å (average, 3.2 Å); most of these structures exhibit the native α-helices. As expected the alternate structures at 500 K are more different from native (mean rmsd range, 4.8–7.8 Å; average, 6.7 Å); most native α-helices are absent. The mean rmsds between all pairs of alternate structures were somewhat less in each case: 1.4–2.5 Å (average, 1.8 Å) at 300 K and 3.0–6.2 Å (average, 4.5 Å) at 500 K. Still the subset of alternate structures for each protein cover a significant amount of conformational space relatively close to the native conformation.

Table 2 summarizes the results of testing the eight empirical models for their ability to discriminate native from alternate structures. Along with the empirical potential energy (or *z* score for the HS model and negative quality score for the BE model) for the native structure, mean and estimated SD values of Δ_i (for the *i*th alternate structure; Eq. **1**) are calculated for the 100 alternate structures generated at each MD temperature. Successful discrimination is defined as a negative Δ_i and the number of unsuccessful discriminations (N^+) at each temperature is also listed in Table 2. Of the models tested, WZS is the only one that is 100% successful at discriminating these alternate structures. LS and HS successfully discriminate

Table 1. Structural characteristics of native and alternate conformations for six test proteins

PDB code	Crystallographic resolution, Å	No. of residues	Temperature, K	R_g(nat), Å	R_g(alt), Å	Mean rmsd, Å	
						Native	Alternate
1crn	1.50	46	300	9.61	9.5 (2)	4.2 (5)	1.9 (6)
			500		10.3 (7)	7.8 (8)	6.2 (23)
2ovo	1.50	56	300	10.1	9.4 (2)	3.5 (3)	2.5 (11)
			500		9.7 (4)	7.7 (9)	4.7 (13)
4pti	1.50	58	300	10.6	10.0 (1)	2.9 (2)	1.5 (5)
			500		9.8 (2)	7.6 (12)	4.2 (12)
1sn3	1.80	65	300	10.5	9.8 (1)	3.2 (5)	1.8 (6)
			500		9.9 (2)	5.7 (10)	4.2 (11)
1hoe	2.00	74	300	11.1	10.3 (1)	2.8 (2)	1.4 (3)
			500		10.7 (2)	6.5 (8)	4.5 (11)
1ubq	1.80	76	300	11.4	10.6 (1)	2.9 (2)	1.4 (4)
			500		10.6 (2)	4.8 (7)	3.0 (7)

PDB, Protein Data Bank; R_g(nat), radius of gyration of the native structure; R_g(alt), mean radius of gyration of the alternate structures; mean rmsd native, mean of rmsd between each alternate structure and the native structure; mean rmsd alternate, mean of rmsd between each pair of alternate structures. The estimated SDs are given in parentheses as deviation in last digit(s) of mean.

all 200 alternate structures for five of six proteins and GS and BE do so for three of six. The few misrecognized structures for these models cannot be considered part of the native ensemble of structures since some of them have relatively large rmsds from native. For example, the 16 1hoe structures misrecognized by LS have rmsds ranging from 2.94 to 3.83 Å. The four structures misrecognized by GS have rmsds ranging from 2.91 to 4.64 Å. MJ succeeds in discriminating all alternate structures only for 1crn, whereas AM and MC fail to discriminate for all six proteins (although MC shows ≈95% discrimination for 4pti and 1hoe).

The means and ranges of the Δ_i values are summarized in Fig. 1, in which results for the alternate structures from a single protein are compared. Examination of Fig. 1 shows that models LS and WZS are the most discriminating, whereas AM, MC, and MJ are least discriminating. This figure presents a clear indication that the WZS model successfully discriminates all the alternate structures of all six proteins. This ability is

Table 2. Results of testing empirical energy models for discrimination of alternative structures

Model	Parameter	Value					
		1crn	2ovo	4pti	1sn3	1hoe	1ubq
AM	E^{native}	−429.95	−614.05	−629.55	−789.53	−829.39	−1005.54
	$\langle\Delta_i\rangle$ (300)	0.06 (2)	0.20 (2)	0.18 (2)	0.13 (1)	0.16 (2)	0.17 (1)
	N^+ (300)	98	100	100	100	100	100
	$\langle\Delta_i\rangle$ (500)	−0.13 (4)	0.15 (3)	0.13 (2)	0.11 (3)	0.12 (3)	0.12 (2)
	N^+ (500)	0	100	100	100	100	100
GS	E^{native}	−7.64	−20.37	−15.54	−19.86	−10.68	−27.59
	$\langle\Delta_i\rangle$ (300)	−0.45 (19)	−0.56 (13)	−0.46 (11)	−0.28 (9)	−0.54 (16)	−0.27 (11)
	N^+ (300)	1	0	0	0	0	2
	$\langle\Delta_i\rangle$ (500)	−0.80 (13)	−0.83 (7)	−0.63 (9)	−0.69 (9)	−0.56 (17)	−0.45 (14)
	N^+ (500)	0	0	0	0	1	0
MC	E^{native}	−119.73	−143.35	−232.28	−157.61	−227.81	−225.81
	$\langle\Delta_i\rangle$ (300)	0.12 (15)	−0.07 (18)	−0.13 (10)	−0.16 (18)	−0.20 (13)	−0.16 (15)
	N^+ (300)	78	36	10	21	8	15
	$\langle\Delta_i\rangle$ (500)	−0.13 (26)	0.00 (19)	−0.36 (14)	−0.38 (26)	−0.40 (15)	−0.29 (16)
	N^+ (500)	33	55	0	9	1	2
BE	$-QS^{native}$	−12.02	−17.75	−18.82	−27.24	−28.66	−33.66
	$\langle\Delta_i\rangle$ (300)	−0.35 (13)	−0.17 (13)	−0.19 (12)	−0.07 (8)	−0.49 (8)	−0.21 (6)
	N^+ (300)	0	8	8	20	0	0
	$\langle\Delta_i\rangle$ (500)	−0.78 (12)	−0.81 (11)	−0.40 (15)	−0.38 (13)	−0.51 (12)	−0.30 (10)
	N^+ (500)	0	0	1	0	0	0
MJ	E^{native}	−257.03	−297.25	−299.94	−273.67	−403.91	−449.55
	$\langle\Delta_i\rangle$ (300)	−0.17 (5)	0.04 (4)	0.03 (4)	0.18 (3)	0.07 (3)	0.03 (4)
	N^+ (300)	0	86	75	100	99	76
	$\langle\Delta_i\rangle$ (500)	−0.49 (8)	−0.27 (5)	−0.10 (7)	0.02 (8)	−0.10 (5)	−0.05 (5)
	N^+ (500)	0	0	8	51	1	17
LS	E^{native}	5.17	−12.92	−34.21	−19.17	15.03	−39.87
	$\langle\Delta_i\rangle$ (300)	−0.85 (1)	−1.36 (39)	−0.81 (16)	−0.99 (32)	−0.21 (18)	−0.53 (16)
	N^+ (300)	0	0	0	0	16	0
	$\langle\Delta_i\rangle$ (500)	−0.89 (1)	−1.19 (6)	−1.76 (20)	−1.42 (14)	−0.59 (12)	−1.01 (27)
	N^+ (500)	0	0	0	0	0	0
HS	z^{native}	−5.57	−5.27	−7.08	−5.95	−5.29	−9.43
	$\langle\Delta_i\rangle$ (300)	−0.25 (8)	−0.04 (8)	−0.25 (5)	−0.22 (8)	−0.38 (7)	−0.17 (6)
	N^+ (300)	0	31	0	0	0	0
	$\langle\Delta_i\rangle$ (500)	−0.72 (12)	−0.90 (16)	−0.60 (10)	−0.66 (16)	−0.63 (13)	−0.32 (11)
	N^+ (500)	0	0	0	0	0	0
WZS1	E^{native}	3.20	−8.21	−12.22	−8.39	−5.34	−27.12
	$\langle\Delta_i\rangle$ (300)	−0.56 (7)	−1.40 (14)	−0.91 (12)	−1.12 (16)	−1.81 (17)	−0.67 (4)
	N^+ (300)	0	0	0	0	0	0
	$\langle\Delta_i\rangle$ (500)	−0.71 (5)	−1.75 (12)	−1.29 (26)	−1.71 (14)	−1.67 (24)	−0.77 (8)
	N^+ (500)	0	0	0	0	0	0
WZS2	E^{native}	−3.05	−13.31	−17.56	−24.35	−16.18	−30.85
	$\langle\Delta_i\rangle$ (300)	−1.51 (37)	−0.89 (11)	−0.58 (9)	−0.56 (8)	−0.81 (12)	−0.56 (4)
	N^+ (300)	0	0	0	0	0	0
	$\langle\Delta_i\rangle$ (500)	−1.43 (14)	−1.37 (21)	−0.84 (22)	−1.05 (15)	−0.93 (22)	−0.61 (6)
	N^+ (500)	0	0	0	0	0	0
WZS3	E^{native}	−3.76	−16.38	−21.48	−30.14	−19.77	−37.59
	$\langle\Delta_i\rangle$ (300)	−1.50 (37)	−0.89 (11)	−0.58 (9)	−0.56 (8)	−0.81 (12)	−0.56 (4)
	N^+ (300)	0	0	0	0	0	0
	$\langle\Delta_i\rangle$ (500)	−1.44 (15)	−1.36 (21)	−0.83 (22)	−1.04 (15)	−0.92 (22)	−0.61 (6)
	N^+ (500)	0	0	0	0	0	0

Models and proteins are defined in text. E^{native} (z^{native}, $-QS^{native}$) is the empirical energy in kcal/mol (z score, negative quality score) of the native structure; $\langle\Delta_i\rangle$ is the mean scaled difference in energy or score (Eq. **1**) for structures at 300 or 500 K with the estimated SD in last digit(s) in parentheses; N^+ is the number of unsuccessfully discriminated alternate structures (for which $\Delta_i > 0$) for structures at 300 or 500 K. In BE, positive quality scores (QS) from Bowie *et al.* (13) were compared as negative numbers since higher QS implies lower energy. In HS, z scores were used in place of energies (27).

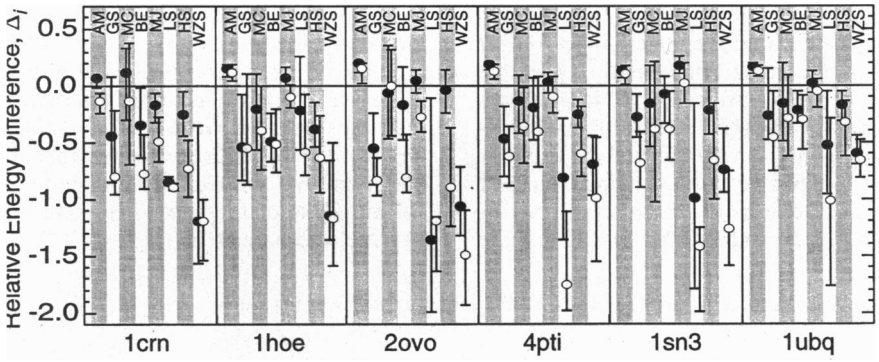

FIG. 1. Mean relative energy differences (Δ_i, Eq. 1) and their ranges for the eight empirical energy models tested (indicated across the top) on alternate structures at 300 K (●) and 500 K (○) of each of six proteins (indicated along the bottom). Every group of 100 alternate structures is represented by a symbol at the mean Δ_i and by bars delimiting the range of Δ_i. The WZS model is the only one for which all $\Delta_i < 0$ for all proteins tested.

independent of which subset of alternate structures are chosen as the training set (cf., WZS1, WZS2, and WZS3 in Table 2).

DISCUSSION

A potential energy model can be deemed successful in thermodynamically discriminating nonnative from native globular protein structures only if it is able to recognize alternate structures distributed throughout conformational space. Alternate protein structures that are conformationally distant from the native structure are expected to be most easily recognized as nonnative, especially if they are noncompact. Thus, an algorithm for generating alternate structures for globular proteins that are both compact and relatively close to native in conformational space is required. The MD method described herein succeeds in generating such alternate structures (Table 1). These alternate structures can also be "tuned" to be within a given rmsd from native by selecting the MD temperature and sampling frequency.

Tests of empirical potential energy models using these alternate structures are found to be significantly more stringent than those based on alternate structures generated by the SRS algorithm (24, 25), which usually generates noncompact alternate structures conformationally distant from the native (rmsds of ≥7 Å) and lacking realistic side chains, effectively emphasizing backbone dihedral angle preferences. The MC empirical energy model successfully discriminated the native structures of 1crn, 2ovo, 4pti, 1sn3, and 1hoe from 13,895, 12,896, 12,701, 12,031, and 11,198 alternate SRS-derived structures, respectively (23), yet failed to discriminate 111, 91, 10, 30, and 9 out of 200 alternate MD-derived structures from each of these proteins (Table 2). The most recent HS model successfully discriminated the native structure of 2ovo from thousands of SRS-derived structures (25) but failed for 31 MD-derived structures (Table 2). Given the enormity of conformational space for even small proteins such as these, any failure to recognize a nonnative structure of any protein from small test sets of alternate structures such as those represented in Table 2 calls into question the utility of the model. The only empirical model in Table 2 to pass this stringent test is WZS.

The AM force field is incapable of recognizing nonnative structures; except for 1crn, AM finds that the alternate structures are on average 10–20% *more stable* than native! The current AM (and other) molecular mechanics force fields are most useful in refining structures already near the conformational minimum. Our results show that AM is incapable of driving a compact alternate structure toward the native structure.

The HS and MJ models derive interresidue energy parameters from a statistical analysis of the known structures of similar sets of proteins, yet HS is much more successful at discriminating alternate from native structures (Table 2 and Fig. 1). The major difference between these models that we believe accounts for this is the continuous nature of the mean force potential energy function in the HS model. MJ uses an all-or-none approach in which interresidue interactions are considered only when the interresidue distance is below some threshold value. The BE and GS models employ the most sophisticated empirical energy functions of the models tested. The BE model includes 18 empirical parameters for the local secondary structure and exposure preferences of each amino acid. The GS model incorporates nearly all existing empirical energy terms, including the Ramachandran potential, hydrogen-bond and rotamer energies, local side chain orientational coupling, one-body centrosymmetric burial potentials, a contact energy, and even multibody side-chain packing interaction energies. However, as seen in Fig. 1, the WZS model, with only 14 empirical parameters, out performs both the BE and GS models. The excellent performance exhibited by the WZS and LS models suggests that the solvation effect dominates the stabilization of native structure in globular proteins (30, 47).

The method used to derive the parameter values for an empirical energy function is as important as the energy function itself. In most models (e.g., MJ, HS, LS, BE, and GS), the parameter values are derived from a statistical analysis of protein crystal structure data. The quantity and quality of available structural data, therefore, put a limitation on derivation of these parameter values. The WZS model requires only a few high-quality structures and enough alternate MD-derived structures can be generated to train the 14 parameters. The MC contact model also uses a learning algorithm to set the energy function parameter values; its poor performance in our tests (Table 2 and Fig. 1) may arise from a poorly chosen training set. The combination of the atomic solvation model and the learning algorithm based on MD-derived compact alternate structures contribute to the ability of the WZS model to successfully discriminate nonnative structures.

In summary, our MD method for generating compact alternate protein structures allows us to perform more stringent tests of the ability of empirical potential energy models to discriminate them from the native structure. In fact, this method exploits the *inability* of the AM molecular mechanics force field to distinguish such nearby local conformational minima on the basis of conformational energy alone (Table 2 and Fig. 1). We have used these compact alternate structures to develop and test any atomic solvation model and found this model to be highly discriminating against nonnative structures.

The help of Dr. J. E. Wampler in reading this manuscript is gratefully acknowledged, Ms. Jane Lu and Dr. Amy Qiu are thanked for assistance in data entry. We acknowledge Drs. M. J. Sippl and M. Jaritz for providing the results for the HS model, Dr. A. Godzik for

providing the results for the GS model, and Dr. D. Eisenberg for providing the three-dimensional-profile program used to generate the results for the BE model. The University of Georgia Computing and Networking Services provided central processing unit time on an IBM RISC/6000 cluster for the computations.

1. Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S. & Weiner, P. (1984) *J. Am. Chem. Soc.* **106,** 765–784.
2. Weiner, P. K., Kollman, P. A., Nguyen, D. T. & Case, D. A. (1986) *J. Comput. Chem.* **7,** 230–252.
3. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983) *J. Comput. Chem.* **4,** 187–217.
4. Novotny, J., Bruccoleri, R. & Karplus, M. (1984) *J. Mol. Biol.* **177,** 787–818.
5. Novotny, J., Rashin, A. A. & Bruccoleri, R. E. (1988) *Proteins Struct. Funct. Genet.* **4,** 19–30.
6. Sikorski, A. & Skolnick, J. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 2668–2672.
7. Skolnick, J. & Kolinski, A. (1990) *Science* **250,** 1121–1125.
8. Skolnick, J. & Kolinski, A. (1991) *J. Mol. Biol.* **221,** 499–531.
9. Shakhonovich, E., Farztdinov, G., Gutin, A. M. & Karplus, M. (1991) *Phys. Rev. Lett.* **67,** 1665–1668.
10. Hinds, D. A. & Levitt, M. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 2536–2540.
11. Andrej, S., Shaknovich, E. & Karplus, M. (1994) *J. Mol. Biol.* **235,** 1614–1636.
12. Covel, D. G. (1994) *J. Mol. Biol.* **235,** 1032–1043.
13. Bowie, J. U., Lüthy, R. & Eisenberg, D. (1991) *Science* **253,** 164–170.
14. Jones, D. T., Taylor, W. R. & Thornton, J. M. (1992) *Nature (London)* **358,** 86–89.
15. Fetrow, J. S. & Fetrow, S. H. (1993) *Biotechnology* **11,** 479–484.
16. Godzik, A. & Skolnick, J. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 98–102.
17. Godzik, A., Kolinski, A. & Skolnick, J. (1992) *J. Mol. Biol.* **227,** 227–238.
18. Ouzounis, C., Sander, C., Scharf, M. & Schneider, R. (1993) *J. Mol. Biol.* **232,** 805–825.
19. Sippl, M. J. & Weitckus, S. (1992) *Proteins Struct. Funct. Genet.* **13,** 258–271.
20. Sippl, M. J. (1993) *Proteins Struct. Funct. Genet.* **17,** 355–362.
21. Tanaka, S. & Scheraga, H. A. (1976) *Macromolecules* **9,** 945–950.
22. Miyazawa, S. & Jernigan, R. (1985) *Macromolecules* **18,** 534–552.
23. Maiorov, V. N. & Crippen, G. M. (1992) *J. Mol. Biol.* **227,** 876–888.
24. Kocher, J.-P. A., Rooman, M. J. & Wodak, S. J. (1994) *J. Mol. Biol.* **235,** 1598–1613.
25. Hendlich, M., Lackner, P., Weitchus, S., Floeckner, H., Froschauer, R., Gottsbacher, K., Casari, G. & Sippl, M. J. (1990) *J. Mol. Biol.* **216,** 167–180.
26. Sippl, M. J. (1990) *J. Mol. Biol.* **213,** 859–883.
27. Sippl, M. J. (1993) *J. Comput. Aided Mol. Des.* **7,** 473–501.
28. Bryant, S. H. & Lawrence, C. E. (1993) *Proteins Struct. Funct. Genet.* **16,** 92–112.
29. Ramachandran, G. & Sasisekharan, V. (1968) *Adv. Protein Chem.* **23,** 283–437.
30. Eisenberg, D. & McLachlan, A. D. (1986) *Nature (London)* **319,** 199–203.
31. Ooi, T., Oobatake, M., Nemethy, G. & Scheraga, H. A. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 3086–3090.
32. Lüthy, R., Bowie, J. U. & Eisenberg, D. (1992) *Nature (London)* **356,** 83–85.
33. Rooman, M. J., Kocher, J.-P. A. & Wodak, S. J. (1991) *J. Mol. Biol.* **221,** 961–979.
34. Rooman, M. J., Kocher, J.-P. A. & Wodak, S. J. (1992) *Biochemistry* **31,** 10226–10238.
35. Wilmanns, M. & Eisenberg, D. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 1379–1383.
36. Kang, H. S., Kurochkina, N. A. & Lee, B. (1993) *J. Mol. Biol.* **229,** 448–460.
37. Goldstein, R. A., Luthey-Schulten, Z. A. & Wolynes, P. G. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 9029–9033.
38. Goldstein, R. A., Luthey-Schulten, Z. A. & Wolynes, P. G. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 4918–4922.
39. Teeter, M. M. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 6014–6018.
40. Bode, W., Epp, O., Huber, R., Laskowski, M. & Ardelt, W. (1985) *Eur. J. Biochem.* **147,** 387–395.
41. Marquart, M., Walter, J., Deisenhofer, J., Bode, W. & Huber, R. (1983) *Acta Crystallogr. B* **39,** 480–490.
42. Almassy, R. J., Fontecilla, J. C. & Suddath, F. L. (1983) *J. Mol. Biol.* **170,** 497–527.
43. Pflugrath, J. W., Wiegand, G., Huber, R. & Vertesy, L. (1986) *J. Mol. Biol.* **189,** 383–386.
44. Vokay-Kumar, S., Bugg, C. E. & Cook, W. J. (1987) *J. Mol. Biol.* **194,** 531–544.
45. Kabsch, W. (1976) *Acta Crystallogr. A* **32,** 922–923.
46. Kabsch, W. (1978) *Acta Crystallogr. A* **34,** 827–828.
47. Holm, L. & Sander, C. (1992) *J. Mol. Biol.* **225,** 93–105.
48. Wesson, L. & Eisenberg, D. (1992) *Protein Sci.* **1,** 227–235.
49. Le Grand, S. M. & Merz, K. M. (1993) *J. Comput. Chem.* **14,** 349–352.
50. Rumelhart, D. E., Hinton, G. E. & Williams, R. J. (1986) *Nature (London)* **323,** 533–536.