J. Mol. Biol. (1997) **266**, 831–846

JMB

Factors Affecting the Ability of Energy Functions to Discriminate Correct from Incorrect Folds

Britt H. Park, Enoch S. Huang* and Michael Levitt

Beckman Laboratories for Structural Biology, Department of Structural Biology, Stanford University of Medicine Stanford, CA 94305-5400, USA

**Corresponding author*

Eighteen low and medium resolution empirical energy functions were tested for their ability to distinguish correct from incorrect folds from three test sets of decoy protein conformations. The energy functions included 13 pairwise potentials of mean force, covering a wide range of functional forms and methods of parameterization, four potentials that attempt to detect properly formed hydrophobic cores, and one environment-based potential. The first of the three test sets consists of large ensembles of plausible conformations for eight small proteins, all of which have correct native secondary structure and are reasonably compact. The second is the set of all subconformations in a database of known protein structures applied to the sequences in that database (ungapped threading). The third is a set of ensembles of 1000 conformations each for seven small proteins taken from molecular dynamics simulations at 298 K and 498 K. Our results show that there are functions effective for each challenge set; moreover, success in one test is no guarantee of success in another. We examine the factors that seem to be important for accurate discrimination of correct structures in each of the test sets, and note that extremely simple functions are often as effective as more complex functions.

Keywords: protein folding; energy function; reduced representation; fold recognition; threading

Introduction

Protein structure prediction has remained elusive because it requires an energy function that can distinguish the correct conformation from the astronomically large number of possibilities (see reviews by Wodak & Rooman, 1993; Sippl, 1995). Such an effective energy function is the common prerequisite for all theoretical approaches to protein folding. Energy functions are often used to drive the conformational changes of a polypeptide chain as it folds through phase space (Wilson & Doniach, 1989; Covell, 1992, 1994; Sun, 1993; Bowie & Eisenberg, 1994; Dandekar & Argos, 1994, 1996; Kolinski & Skolnick, 1994; Wallqvist & Ullner, 1994; Vieth *et al*., 1994, 1995; Monge *et al*., 1995; Mumenthaler & Braun, 1995; Srinivasan & Rose, 1995; Sun *et al*., 1995). Alternatively, they are employed to discriminate amongst candidate (or ''decoy'') folds generated by methods that are independent (or semi-independent) of the energy function. For instance, many investigators elect to use large sets of decoy folds constructed by ungapped threading methods (Hendlich *et al*., 1990; Maiorov & Crippen, 1992; Bryant & Lawrence, 1993; Kocher *et al*., 1994; Sippl, 1995; Huang *et al*., 1995). The resulting structures tend to be non-compact and quite dissimilar to the actual target fold and hence do not present a serious challenge to many discrimination functions. Using functions to align the sequences upon existing structures (i.e. ''inverse protein folding'') can yield candidates that more closely resemble the native structure (Bowie *et al*., 1991; Ouzounis *et al*., 1993; Sippl & Weitckus, 1992; Jones *et al*., 1992; Godzik *et al*., 1992; Bryant & Lawrence, 1993; Wilmanns & Eisenberg, 1993; Fischer & Eisenberg, 1996). Other studies build and evaluate decoys that differ only slightly from the native using molecular dynamics (MD) simulations of crystal structures (Wang *et al*., 1995a,b; Huang *et al*., 1996). Simulations at room temperature can easily produce thousands of structures that only differ slightly from the crystal structure (Huang *et al*., 1996). A drawback to room temperature simulation methods is that the conformational space they explore is quite limited. High temperature simulations sample more of phase

Abbreviations used: MD, molecular dynamics; RMS, root-mean-square.

0022–2836/97/090831–16 $25.00/0/mb960809

space but yield structures that compromise the compactness of the fold, the integrity of secondary structure elements, and the close packing of the side-chains (Huang *et al*., 1996). Finally, candidate folds may be generated by exhaustive searches either using lattice (Hinds & Levitt, 1992, 1994) or off-lattice representations (Cohen *et al*., 1979; Park & Levitt, 1995, 1996; Yue & Dill, 1996). Lattice representations must sacrifice structural features such as secondary structure elements or else employ many (>12) states per residue. However, our off-lattice model, in which secondary structure elements are held fixed and all conformations of selected loop residues are explored, can generate conformations within a 2 Å coordinate root-mean squared-deviation (RMS) of the native structure using only four discrete dihedral angle states (Park & Levitt, 1995).

Clearly, a good energy function lies at the heart of all structure prediction methods: (1) as a driving function for folding polypeptides from an extended or random state, or (2) as a scoring scheme to distinguish native (or native-like) conformations from a pool of candidates generated independently of the function itself; or (3) a way of aligning sequence upon existing structures to detect native-like matches. It is not clear whether energy functions that are good for one of these tasks are also good for the others. Does there exist a single function that is universally effective? If no single method is effective against all types of decoys, i.e. if different challenges require particular types of energy functions, then we must elucidate the underlying common features amongst the functions that make them specific for the different problems.

In previous work we have examined the discrimination power of some commonly used types of energy functions using large ensembles of plausible folds with correct secondary structure (Park & Levitt, 1996). We have also investigated the effectiveness of a simple hydrophobic fitness function to identify correct folds from ensembles generated by ungapped threading and molecular dynamics simulations (Huang *et al*., 1995, 1996). In this work, we expand our study of energy functions that distinguish native and near-native structures from non-native structures (i.e. the second class of problems listed above). In addition to introducing both new potentials and variants of the potentials tested in our earlier work (Huang *et al*., 1995, 1996; Park & Levitt, 1996), we also apply each function to the entire spectrum of decoy sets (ungapped threading, MD simulation, and exhaustive enumeration).

Success (or failure) in each set provides useful information about how each function might be best used. For instance, models produced by inverse folding result from optimizing the alignment of sequence upon the template using the energy function and search strategy such as the ''branch and bound'' algorithm (Lathrop & Smith, 1995) or some variant of a dynamic programming algorithm (Needleman & Wunsch, 1970; Taylor & Orengo, 1989). Therefore, if the function is defeated by decoys that are unoptimized in sequence-template alignment (i.e. by ungapped threading), it is unlikely to be useful for inverse folding. Functions that perform well against near-native decoys generated by MD are likely to be useful for minimization in continuous space, perhaps in the final stages of building a model by *ab initio* means or selecting among models generated by alignment upon known homologues. Finally, a function that screens effectively against exhaustively generated folds by our four-state procedure selects folds that are compact, globular, and exhibit proper tertiary arrangement. These are qualities desirable in a function integral in any *ab initio* approach.

This broad-reaching study aims to answer some basic questions regarding energy functions. By modifying the potentials discussed in earlier work (Huang *et al*., 1995, 1996; Park & Levitt, 1996), we can assess the effect of changing the functional forms on the discrimination ability in the various challenges. Quantitative measures of the performance of each function on each challenge set are provided. The similarities in the different potentials, evaluated by the correlations in performance between pairs of energy functions, will also be discussed. We will show how the competence of the energy functions varies with the type of challenge and discuss some of the underlying causes of the observed differential performance.

Results

The procedure that we followed in this study was simple. We tested 18 reduced representation empirical energy functions, most of which are variants of functions typically used for *ab initio* protein structure prediction or sequence-structure matching on three different sets of possible conformations. The energy functions examined are described in Methods and summarized in Table 1.

The functions we examined can be classified by two broad characteristics. The first is the reference state necessary for the generation of parameters in a statistical-mechanical framework. The Contact(MJ), VdW(MJ), VdW(MJ)12, and VdW(MJ)4 functions all have the unfolded state as reference. The Contact(HL), VdW(HL), Histogram, VdW(HL)12, VdW(HL)4, Shell, Shell(top), Shellm, and Shell(top)m, HF(stat) and HF(stat)m functions all have the compact state as reference. The Surface, HF, and HF(sm) functions do not have a reference state since they are not derived statistically from the database of known structures. The second broad classification of the energy functions is that of distance dependence. Strictly speaking, all of the energy functions are distance-dependent: even for simple contact-based functions a pair of residues typically has an energy of e_{ij} if they are closer than some cutoff distance and 0 if they are not. For the purposes of the current study, however, we speak of functions that have only an on/off dependency

Table 1. Energy functions examined

Energy function	Description
Contact(MJ)	Residue-residue contact function. The reference state is the unfolded polypeptide chain, and energy parameters are calculated from contact frequencies determined by atom-atom contacts in database of known structures. The α-carbon backbone of each residue is treated as a distinct interacting center in addition to a side-chain interacting center
Contact(HL)	Residue-residue contact function. Identical to Contact(MJ), except the reference state is the compact state
VdW(MJ)	Distance-dependent residue-residue potential. The energy parameters for the Contact(MJ) function are fit to a function of the form A/r^8-B/r^4
VdW(HL)	Distance-dependent residue-residue potential. Identical to VdW(MJ) except based on Contact(HL) parameters
Surface	Crude estimate of the exposed hydrophobic surface area using the approximate solvent accessibility algorithm of Wodak & Janin (1980)
Histogram	Pairwise potential of mean force which calculates interaction energies based on residue-residue separations in space and in sequence (Hendlich *et al.*, 1990)
VdW(MJ)12	Distance-dependent residue-residue potential. Identical to VdW(MJ) except the functional form is A/r^{12}-B/r^6
VdW(MJ)4	Distance-dependent residue-residue potential. Identical to VdW(MJ) except the functional form is A/r^4-B/r^2
VdW(HL)12	Distance-dependent residue-residue potential. Identical to VdW(MJ)12 except based on Contact (HL) parameters
VdW(HL)4	Distance-dependent residue-residue potential. Identical to VdW(MJ)4 except based on Contact(HL) parameters
Shell	Residue-residue contact function. Parameters are derived from the frequency of residue-residue contacts determined by a simple distance cutoff of 7 Å
Shell(top)	Residue-residue contact potential. Like the Shell function but residue-residue contact frequencies are normalized by relative likelihood of contacts between residues separated by different numbers of amino acids in sequence
Shellm	Residue-residue contact potential. Like the Shell function, but uses five different sets of parameters, each of which are calculated using different distance cutoffs (6, 7, 8, 9 and 10 Å)
Shell(top)m	Residue-residue contact potential. Like the Shell(top) function, but uses five different sets of parameters
HF	Hydrophobic fitness function. An energy function without parameters that scores for the formation of true hydrophobic cores (Huang *et al.*, 1995)
HF(sm)	A variant of the HF potential that uses a sigmoidal function for counting residue contacts rather than a step function
HF(stat)	Statistically derived hydrophobic fitness function. Score is based on a potential of mean force for the number of residues within 7 Å of each residue
HF(stat)m	Like HF(stat) except five different distance cutoff ranges (6,7,8,9 and 10 Å) are used

as non-distance-dependent. The Contact(MJ), Contact(HL), Shell, Shell(top), and HF functions are all therefore non-distance-dependent. All the other functions are distance-dependent, either because they have a continuous functional form or because they have distinct energies for different inter-residue distances.

The three test sets of conformations used in this study are described in Methods and summarized in Table 2. The first test set consists of ensembles of plausible conformations generated exhaustively for eight small proteins. All conformations in these ensembles have correct secondary structure, and are reasonably compact. They have RMS deviations ranging from ~2 Å to 15 Å from the appropriate X-ray structure. The second test set consists of possible substructures from a database of known X-ray structures applied to all sequences in that database, the ungapped threading test. The RMS error for this set ranges from ~8 Å to over 30 Å. The final test set consists of ensembles of 1000 conformations each generated by molecular dynamics simulation at 298 K and 498K for five small proteins. The RMS deviations fall between 0 and ~3 Å for the former and ~10 Å for the latter.

Figure 1 shows the distribution of RMS deviations for typical examples of each of our test sets for

Table 2. Test sets used

Test set	Description
Ensembles of plausible conformations	Large ensembles of from 30,000 to 200,000 conformations generated by exhaustive exploration of loops while preserving native secondary structure for 8 small proteins (Park & Levitt, 1996)
Ungapped threading	103 protein sequences are superimposed without any gaps on all possible contiguous subconformations of the same set of 103 proteins. The number of conformations available to each protein ranges from 1 to about 20,000 (Huang *et al.*, 1995)
Molecular dynamics ensembles	1000 conformations each for 7 small proteins taken from 1 ns trajectories generated by molecular dynamics simulation in solution at 298 K and 498 K (Huang *et al.*, 1996)

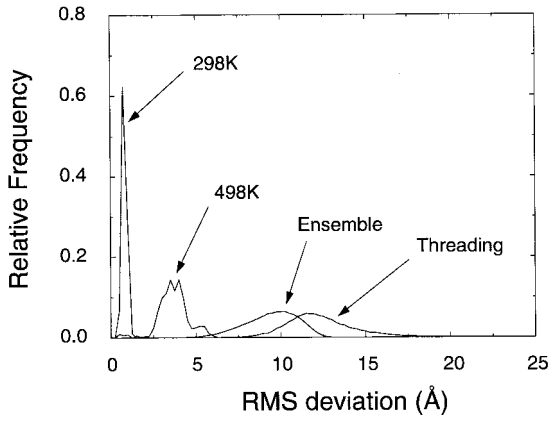

Figure 1. Distribution of RMS deviations from the X-ray structure for different test sets for the protein 1r69.

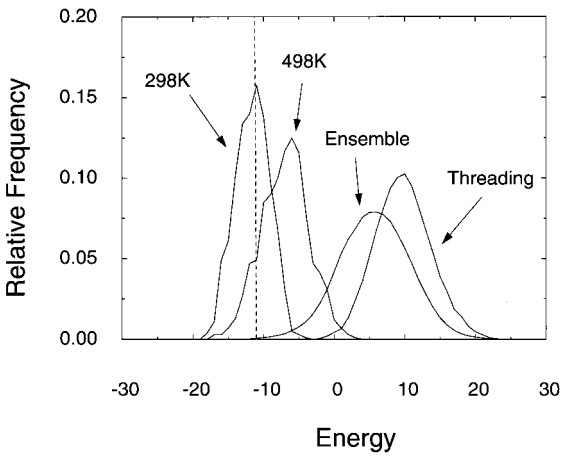

Figure 2. Distribution of Shellm energies for the different test sets for the protein 1r69.

1r69. Figure 2 shows their energy distributions for a representative function (Shellm). There is significant overlap between the different test sets, both in RMS deviation and energy. As the sets become less like the X-ray structure (larger RMS), the energy also tends to be less favorable; this behavior does, however, depend on which energy function is used. It is also noteworthy that the different ensembles have distinct characteristics. The ungapped threading structures have RMS deviations that overlap those of the ensembles of plausible conformations, yet the latter are mostly compact while the former are not. The 498 K simulations provide structures that are somewhat compact but have distorted secondary structure.

We proceed now to describe the results for each of these test sets in turn, and then concentrate on the different functions' overall strengths and weaknesses.

Ensembles of plausible conformations

There are two questions that we can ask about energy functions using our ensembles of plausible conformations: how well can they identify the correct structure, and how well can they identify nearly correct structures? These ensembles are particularly useful for the latter question because they contain many conformations within 4 Å RMS of the correct structures and several within 2 Å RMS. Here we refer to structures nearer than 4 Å to the appropriate X-ray structure as native-like.

Identification of X-ray conformations

Table 3 shows the Rank scores, average Q-scores and average Z-scores for X-ray structure identification from our ensembles of plausible conformations. The results for the Contact(MJ), Contact(HL), VdW(MJ), VdW(HL), Surface, and Histogram energy functions were presented by Park & Levitt (1996) but are shown again to facilitate comparisons. Several of the new functions all have average Q-scores ($\langle Q \rangle$) better than the best described previously (the VdW(MJ) function). These include the VdW(MJ)12, VdW(HL)4, Shellm and Shell(top)m functions. In particular, the first two of these rank X-ray structure first in energy as often as not. The VdW(MJ)12 energy function is even comparable to the best combination of energy functions described by Park & Levitt (1996), which had $\langle Q \rangle = 5.10$. It is remarkable how a simple change of functional form improves performance so drastically.

For this test there are also several disappointing functions. The two variants of the HF function, (HF and HF(sm)), the original of which does very well in ungapped threading (Huang *et al.*, 1995; see below), although not the worst functions, are only middling performers. It should be noted, however, that these HF functions stumbled most noticeably for 1sn3 and 4pti, two small, heavily disulfide-bonded proteins. The HF score, which does not recognize disulfide bonds, is not expected to perform well for these two proteins. The Contact(MJ) function, with its average Q-score of 0.46, is marginally better than a random discriminator ($\langle Q \rangle = 0.3$).

In our previous survey of energy functions we concluded that distance-dependent energy functions worked better than non-distance-dependent functions. Indeed the two best functions in this study, VdW(MJ)12 and VdW(HL)4, are distance-dependent. However, we see now that two non-distance-dependent functions, Shell and Shell(top), work quite well. Moreover, the distance-dependent version of the HF function, HF(sm), performs considerably worse than its non-distance-dependent analog, HF, although there are other differences between the two functions besides distance dependence. In any case, smoothing of energy functions is not a certain route to improved performance.

Table 3. X-ray rank from ensembles of plausible conformations

Protein	Contact (MJ)	Contact (HL)	VdW(MJ)	VdW(HL)	Surface	Histogram	VdW(MJ) 12	VdW(MJ) 4	VdW(HL) 12	VdW(HL) 4
4rxn	29,184	10,154	49	324	285	12,277	3	6942	1388	2
4pti	126,900	129,987	286	5074	10,963	1488	4	30,719	4763	12,293
1r69	19,423	78	77	222	2939	2703	1	10,174	810	1
2cro	57,484	478	160	8	19,092	8	1	16,765	157	4
1sn3	55,387	16,067	2	8	31,000	6188	1	7609	61	2
1ctf	54,658	46	2	16	2592	1	1	10,095	118	1
3icb	141,032	234	1327	32	6965	1	3	40,007	201	1
1ubq	22,527	63	1	1	719	7	1	15,374	1	1
$\langle Q \rangle$	0.46	2.19	3.66	3.53	1.55	3.13	4.96	1.01	2.90	4.49
$\langle Z \rangle$	−0.21	−1.85	−3.95	−2.53	−1.78	−2.91	−3.98	−1.27	−2.34	−3.23

Protein	Shell	Shell(top)	Shellm	Shell(top)m	HF	HF (sm)	HF(stat)	HF(stat)m
4rxn	41	16	85	31	253	13,791	1	2
4pti	473	627	246	161	17,811	17,441	105	42
1r69	131	227	158	30	201	7287	1173	4
2cro	84	495	103	85	228	36,340	7532	563
1sn3	610	469	314	370	7464	6153	1828	231
1ctf	9	3	1	1	472	14,292	14	1
3icb	1	1	1	1	42	31,021	1	1
1ubq	2	1	5	1	9	10,458	3	9
$\langle Q \rangle$	3.61	3.63	3.69	3.95	2.60	1.00	3.42	4.08
$\langle Z \rangle$	−3.65	−3.84	−3.46	−3.96	−4.15	−1.31	−1.81	−2.31

Identification of native-like conformations

Table 4 shows the best Rank scores, average *Q*-scores and average *Z*-scores for native-like structure identification from our ensembles of plausible structures. As one would expect, the discrimination of native-like conformations is poorer than the discrimination of X-ray structures for all functions. Especially interesting, however, is the performance of the different functions relative to each other.

Here we find the best performing functions to be the Shell(top) and Shell(top)m functions with average *Q*-scores of 2.21 and 2.10, respectively. Although both these functions are reasonable performers for X-ray structure identification, they are not the best. The VdW(HL)4 function, which excelled at identifying X-ray structures, is still third best with an average *Q*-score of 1.92. However, the VdW(MJ)12 function, which also excelled, slips markedly relative to the others ($\langle Q \rangle = 1.33$).

It is clear from Table 4 that the importance of distance dependence, *per se*, is less obvious for native-like structure identification. The best function, Shell(top), is not distance-dependent, and two other non-distance-dependent functions, Shell and Contact(HL), do reasonably well with average *Q*-scores above 1.5.

Ungapped Threading

Table 5 shows the results for the ungapped threading test. The first row shows the percentage of sequences for which the correct structure is ranked first in energy of all applicable substructures in the database of 103 proteins. The second and third row show the average *Q*-scores and *Z*-scores. Six of the 18 functions can be considered quite successful: the Histogram, Shell, Shell(top), Shellm, Shell(top)m, and HF functions all succeed more than 80% of the time. Indeed, those sequences for which these functions fail are almost all either membrane proteins, polypeptides that are normally part of strongly bound multisubunit proteins, proteins that have large prosthetic groups, or small proteins that depend strongly on disulfide bonds for stability.

The most interesting aspect of these results is how they differ from the results found for our ensembles of plausible conformations. The best function in this test, the Shell(top)m, is reasonably good at identifying X-ray structures from the plausible ensembles, and it was among the best at finding native-like structures. On the other hand, the VdW(MJ)12 and VdW(HL)4 functions, which were superb at X-ray structure identification and reasonably effective at native-like structure identification, are mediocre performers here. The Surface energy function, although no better than the aforementioned VdW functions, is greatly improved relative to its performance in the ensembles of plausible structures.

The average *Q*-score and frequency of success are perfectly rank-correlated for the top six performers. These top performers also have the best (most negative) average *Z*-scores, although the rank correlation is lower. One anomaly is that the HF function has a disproportionately favorable average *Z*-score. The Histogram function, for example, succeeds in identifying correct structures as often as the HF function and yet has

Table 4. Best native-like ranks from ensembles of plausible conformations

Protein	Contact (MJ)	Contact (HL)	VdW(MJ)	VdW(HL)	Surface	Histogram	VdW(MJ) 12	VdW(MJ) 4	VdW(HL) 12	VdW(HL) 4
4rxn	2116	1	15	5	330	1840	100	57	2	2
4pti	3949	364	101	35	768	416	876	1480	445	106
1r69	1270	115	71	135	496	34	34	230	413	31
2cro	195	20	27	129	153	15	15	28	149	9
1sn3	16,239	122	149	3525	1141	7530	198	2614	5786	439
1ctf	24,156	28	98	28	2454	40	81	3,624	97	40
3icb	3538	86	533	2	373	6	16	842	6	1
1ubq	9895	46	18	22	22	182	32	4,100	55	6
⟨*Q*⟩	−0.40	1.52	1.33	1.51	0.57	1.01	1.33	0.37	1.15	1.92
⟨*Z*⟩	0.03	−0.74	−0.98	−0.89	−0.87	−1.27	−0.96	−0.43	−0.73	−1.45

Protein	Shell	Shell(top)	Shellm	Shell(top)m	HF	HF (sm)	HF(stat)	HF(stat)m
4rxn	2	3	74	30	27	55	97	20
4pti	49	31	104	106	2468	10,288	1923	114
1r69	22	16	41	11	40	653	240	125
2cro	16	5	25	9	6	49	439	61
1sn3	47	4	47	11	42	1873	512	942
1ctf	91	9	1	1	2342	4241	5	3
3icb	6	4	1	1	22	547	3	44
1ubq	93	29	88	76	63	3,443	22	2
⟨*Q*⟩	1.78	2.21	1.83	2.10	1.23	0.22	1.21	1.57
⟨*Z*⟩	−1.53	−1.63	−1.54	−1.74	−1.51	−0.59	−0.81	−1.04

an average *Z*-score of −4.05 as compared to −11.31 for the HF function.

Molecular dynamics generated ensembles

Our third test set consists of ensembles of conformations generated by molecular dynamics simulations at 298 K and 498 K for seven small proteins. There are 1000 conformations in each set, one conformation for every picosecond of the one nanosecond simulations.

298 K conformations

The conformations generated by 298 K simulations have mean RMS deviations of 1.72 Å from the starting X-ray structures. One therefore expects that these sets will be a fairly severe test. For the most part this turns out to be true. Several functions barely do better than chance, namely the VdW(HL), Histogram, VdW(HL)12 and HF(stat) energy functions.

Most surprising, however, is the fact that some functions do quite well (Table 6). The Contact(MJ), VdW(MJ), Surface, VdW(MJ)4, HF and HF(sm) functions all have average *Q*-scores better than 1.5. It is particularly interesting to note that the Surface function does best of all. For the ensembles of plausible conformations this function is a mediocre performer. Figure 3 shows a plot of RMS deviation *versus* Surface energy for the 298 K and 498K dynamics structures of ubiquitin (1ubq). In contrast is Figure 4, in which the same plot is shown for the Histogram energy function, a poor performer. The much better performance of the Surface function (Figure 3) is obvious in the striking correlation of

Table 5. Structure identification from ungapped threading

	Contact (MJ)	Contact (HL)	VdW(MJ)	VdW(HL)	Surface	Histogram	VdW(MJ) 12	VdW(MJ) 4	VdW(HL) 12	VdW(HL) 4
% Successes	9	44	25	27	64	82	42	12	8	64
⟨*Q*⟩	1.61	2.57	2.16	1.79	3.12	3.42	2.60	1.77	0.82	3.26
⟨*Z*⟩	−1.88	−3.21	−2.04	−1.96	−3.20	−4.05	−2.74	−1.65	−0.57	−4.50

	Shell	Shell(top)	Shellm	Shell(top)m	HF	HF (sm)	HF(stat)	HF(stat)m
% Successes	86	87	89	90	84	61	45	64
⟨*Q*⟩	3.59	3.63	3.65	3.65	3.56	3.21	2.98	3.27
⟨*Z*⟩	−7.08	−7.48	−7.07	−7.69	−11.31	−4.83	−2.76	−3.04

Table 6. X-ray ranks from 298 K dynamics ensembles

Protein	Contact (MJ)	Contact (HL)	VdW(MJ)	VdW(HL)	Surface	Histogram	VdW(MJ) 12	VdW(MJ) 4	VdW(HL) 12	VdW(HL) 4
1ctf	635	393	380	880	12	128	352	346	778	355
1hdd	157	125	79	636	1	434	229	28	365	135
1pgb	64	126	7	1	1	407	147	16	1	2
1r69	1	15	19	640	3	1001	32	16	424	199
1ubq	1	12	2	341	3	930	129	5	205	233
4icb	171	438	1	333	1	397	330	1	925	14
8lyz	1	1	1	1001	1	1	41	1	1001	5
⟨*Q*⟩	1.71	1.33	2.01	0.63	2.71	0.72	0.88	1.99	0.66	1.36
⟨*Z*⟩	−2.13	−1.80	−2.63	0.16	−3.07	−0.20	−1.08	−2.91	0.46	−1.49

Protein	Shell	Shell(top)	Shellm	Shell(top)m	HF	HF(sm)	HF(stat)	HF(stat)m
1ctf	806	858	955	748	748	27	654	418
1hdd	88	468	451	124	2	1	974	184
1pgb	420	810	319	438	173	191	5	1
1r69	194	472	271	390	88	1	346	7
1ubq	108	621	692	655	40	88	396	766
4icb	2	23	39	6	1	1	259	39
8lyz	1	1	1	1	1	1	962	828
⟨*Q*⟩	1.27	0.81	0.85	1.03	1.72	2.19	0.57	1.13
⟨*Z*⟩	−1.58	−0.62	−1.22	−1.50	−2.46	−4.18	0.05	−0.85

energy and RMS deviation. Also surprising is the fact that the HF(sm) function comes in second, beating the HF function by 0.47 in average *Q*-score. For the ensembles of plausible conformations and for ungapped threading the HF(sm) function is a distinctly poorer performer than the HF function.

498 K structures

Conformations generated by molecular dynamics at 498 K are much further on average from the starting X-ray conformations than those from the 298 K simulations, with an average RMS deviation from the X-ray structures of 5 Å. Moreover, at this temperature the secondary structure elements are partially unfolded. We would therefore expect that these conformations will be less of a challenge to the energy functions than the 298 K conformations.

Indeed they are. Table 7 shows that six energy functions, Contact (MJ), VdW(MJ), Surface, VdW(MJ)4, HF and HF(sm) all achieved average *Q*-scores of 2.7 or better. Perhaps the most interesting finding with this test set is the number of functions that discriminate extremely poorly. The VdW(HL) and VdW(HL)12 functions are even anti-selective, so to speak. They consistently rate the correct structures as among the highest in energy.

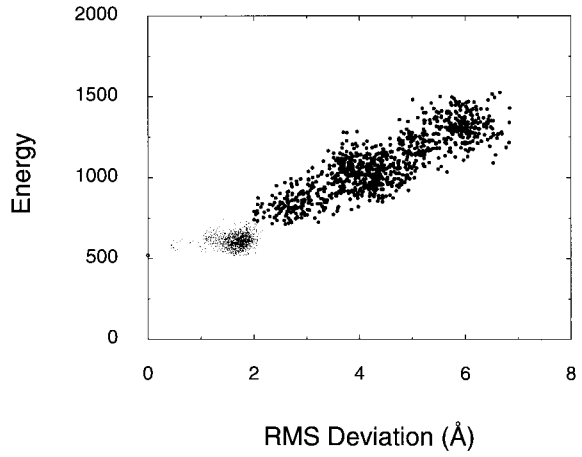

Figure 3. RMS deviation (in Å) from the native structure of ubiquitin (1ubq) as a function of Surface energy for the 298 K (small points) and 498 K (large points) dynamics conformations. The energy of the X-ray structure is shown as a small circle on the *Y* axis.

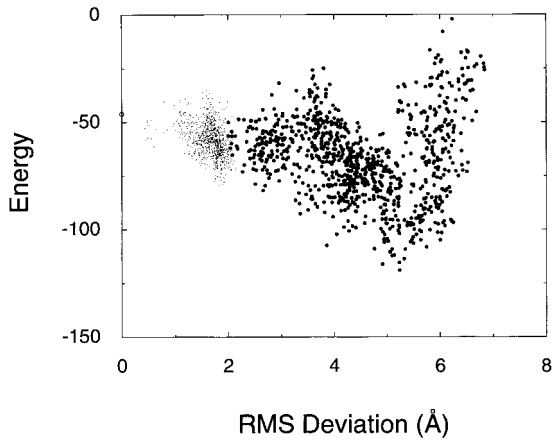

Figure 4. RMS deviation (in Å) from the native structure of ubiquitin (1ubq) as a function of Histogram energy for the 298 K (small points) and 498 K (large points) dynamics conformations. The energy of the X-ray structure is shown as a small circle on the *Y* axis.

These two functions also do worse for the 498 K ensembles than the 298 K ensembles.

Function similarities

The differential performance of the 18 energy functions led us to wonder whether there was some objective way to identify similarities and differences among the energy functions that were not obvious from the definitions of the functions. To be more precise, we noted at the beginning of Results that the reference state used for parameter generation and distance-dependence are two obvious characteristics by which the energy functions can be classified. Are there any other less obvious similarities among the functions?

We have attempted to answer this question, at least partially, by mapping the energy functions to points in two-dimensional space, using the correlation in performance for different test sets between all pairs of functions as a distance metric. The details of this procedure are described in Methods. Figure 5 shows a plot of the positions of the different energy functions in this two-dimensional projection. The distance between two functions in this space corresponds to the correlation in performance between those functions over the different test sets. The closer together two functions are, the more highly correlated their performances.

The most obvious feature in Figure 5 is the clustering of the Histogram function and all the Shell-style functions. *A priori*, one would expect the Shell-type functions to be highly similar. The fact that the Histogram function is so similar to the Shell-type functions is interesting. Perhaps this, together with the fact that the various Shell functions are better than the Histogram function in all our test sets, is an indication that the additional complexity of the Histogram function is unnecessary. The Shell and Shell(top) functions have 210 parameters while the Histogram function has 80,000.

Keeping in mind that the similarities shown in Figure 5 are likely to be suggestive only, it is still interesting to note that the (HL) and (MJ) style functions are segregated. The broken curve shows the boundary. All of the Shell style functions and the Histogram function are also on the same side of the line as the (HL)-derived functions. It is notable that two functions are close to the dividing line: VdW(MJ)12 and VdW(HL)4. These functions are the best performers in the plausible ensemble test.

We can also see two isolated pairs of functions that are also highly similar: Surface-HF(sm) and Contact(MJ)-VdW(MJ)4. The similarity of the Surface and HF(sm) functions is to be expected. They are both fairly simple measures of hydrophobic burial. The other similar pair is difficult to understand. The Contact(MJ) and VdW(MJ)4 functions are hydrophobicity-emphasizing functions (they have the unfolded state as reference), but so are the VdW(MJ) and VdW(MJ)12 functions, yet the latter pair is quite dissimilar.

Discussion

In recent years there has been a plethora of studies reporting empirical energy functions and their performance under particular circumstances (see Introduction for references). Here we have presented

Table 7. X-ray ranks for 498 K dynamics ensembles

Protein	Contact (MJ)	Contact (HL)	VdW(MJ)	VdW(HL)	Surface	Histogram	VdW(MJ) 12	VdW(MJ) 4	VdW(HL) 12	VdW(HL) 4
1ctf	1	18	1	804	1	15	1	1	934	36
1hdd	61	12	8	911	1	218	216	1	938	23
1pgb	2	23	1	240	1	897	10	2	863	14
1r69	1	2	1	730	1	797	1	1	871	2
1ubq	1	1	1	623	1	880	14	1	807	45
4icb	1	10	1	415	1	159	17	1	894	1
8lyz	1	1	1	1001	1	1	1	1	1001	1
⟨Q⟩	2.70	2.28	2.87	0.21	3.00	0.92	2.18	2.96	0.04	2.14
⟨Z⟩	−3.69	−3.18	−3.80	0.85	−3.70	0.53	−2.39	−4.19	1.91	−2.63

Protein	Shell	Shell(top)	Shellm	Shell(top)m	HF	HF(sm)	HF(stat)	HF(stat)m
1ctf	58	84	118	40	27	1	167	74
1hdd	79	271	408	130	1	1	644	13
1pgb	361	658	207	212	2	5	2	2
1r69	40	87	37	18	1	1	55	2
1ubq	57	182	393	129	2	1	41	62
4icb	2	3	6	2	1	1	5	1
8lyz	1	1	1	1	1	1	47	30
⟨Q⟩	1.58	1.30	1.29	1.61	2.71	2.90	1.00	2.02
⟨Z⟩	−2.25	−1.69	−1.68	−2.15	−4.24	−5.64	−1.64	−2.43

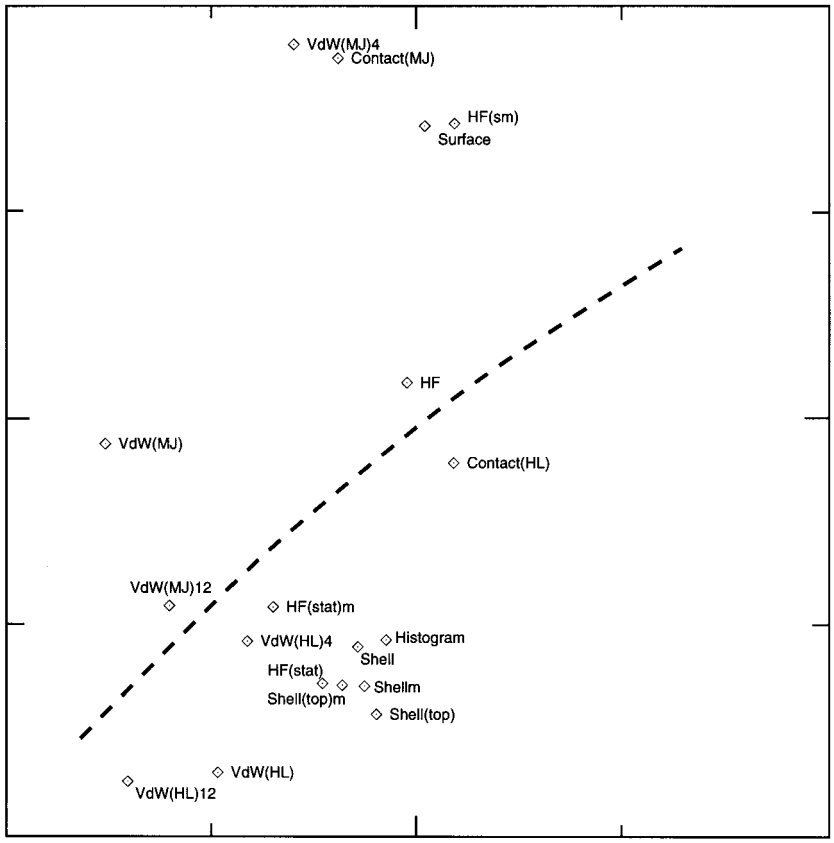

Figure 5. Projection of the correlations of energy function performance onto two dimensions. The broken curve represents a hypothetical dividing line between functions that emphasize hydrophobicity and those that do not. The scale of this plot is the same for both axes but are in all other ways arbitrary.

the most comprehensive examination of different energy function classes to date, using three complementary test sets. The most remarkable finding is that energy functions designed for one purpose do not necessarily work well for different purposes. Functions that we found to be effective at discriminating X-ray structures from ensembles of plausible conformations (similar to the kinds of conformations likely to be found in *ab initio* folding simulations) are found here to be less than ideal for picking out correct sequence-structure matches in ungapped threading tests. We also find that some functions that are effective for the ungapped threading problem, the HF function for instance, are considerably less effective at finding X-ray structures in the ensembles of plausible conformations. The same inconsistency is found for identifying correct structures from the dynamics ensembles. The VdW(HL) function, which is reasonably good at identifying X-ray structures from the ensembles of plausible conformations, is actually counter-selective for finding X-ray structures from the dynamics ensembles.

The lesson to be learned from this fickle behavior is that the whole business of energy function design is extremely sensitive to the targeted problem domain. Different problems require different function characteristics.

What makes functions work (or not work)?

Although it useful to know which kinds of functions work well in different problem domains, it would be much more informative if one could also know why. To do this at least partially we first ask the question, what are the common characteristics of functions that do well for each of the test sets?

For identifying X-ray structures from the ensembles of plausible conformations, the two best functions are VdW(MJ)12 and VdW(HL)4, both of which are distance-dependent and have two interacting centers (corresponding to the C^{α} and side-chain). Provisionally we can say that those two characteristics are important. However, some other distance-dependent functions, VdW(MJ)4 and HF(sm) for example, are poor performers. What, then, is there that distinguishes the VdW(MJ)12 and VdW(HL)4 functions? In our previous study (Park & Levitt, 1996) we found that combining the VdW(MJ) and VdW(HL) functions gave a combined function which performed extremely well (essentially as well as either the VdW(MJ)12 or VdW(HL)4 functions of this study). Our conclusion then was that the VdW(MJ) function overemphasized and the VdW(HL) function underemphasized purely hydrophobic interactions. The balance between the hydrophobic and residue-specific components was critical for function performance. We think that something similar happens when

the functional forms of the VdW functions change. Using the A/r^{12}-B/r^6 functional form for the VdW(MJ)12 function causes interactions at longer distances to be de-emphasized. Using the A/r^4-B/r^2 functional form for the VdW(HL)4 function causes interactions at longer distances to have relatively greater weight. Since the number of interactions between residues increases at long distances, one expects that those interactions will be, on average, less residue-pair specific. Each specific interaction at long range will have a smaller fractional contribution to make to the total energy. A concrete example may make this reasoning clearer. Consider two on/off contact functions, the first of which counts residues as contacting if they are within 6 Å of each other, the second if they are within 10 Å. The average number of residues in contact with any particular residue will be ~2.8 for the first function, the average number of neighboring residues within 6 Å in our database. For the 10 Å cutoff function the average number of contacts for any particular residue is ~14.9. One can see that the residue specificity of interactions will be lower for the 10 Å cutoff since the fractional contribution of any specific contact is only 1/14.9 compared to 1/2.8 for the 6 Å cutoff. Functions that weigh distant interactions more heavily, like the VdW(HL)4 function, will therefore reward general hydrophobic burial more strongly and specific residue-residue interactions more weakly. Conversely, functions that count distant interactions less, like the VdW(MJ)12 function, will weigh residue-specific interactions more strongly and general hydrophobic burial more weakly. Some further weight is lent to this hypothesis by the observation (in Figure 5) that the VdW(HL)4 and VdW(MJ)12 functions are most similar in their performance profile of all VdW(MJ) and VdW(HL) pairs.

In the case of identifying native-like tertiary folds from the ensembles of plausible conformations, different factors are important. Here the value of distance dependence and C^α backbone dependence are not really evident. The VdW(HL)4 function is still a good performer (third best of all), but the VdW(MJ)12 function falls back and is just an average discriminator. The best functions are the Shell(top) and Shell(top)m functions, the latter distance-dependent and the former not. Overall, it appears that additional complexity does help somewhat, as the Shell function was the weakest of the Shell-type functions. However, the Histogram function, which has far more parameters than even the most complex Shell function, underperformed even the simplest Shell function, supporting our hypothesis that additional complexity does not necessarily improve performance.

The fact that C^α interactions do not seem as important for near-native discrimination, compared to X-ray structure identification, may reflect the ability of the VdW style functions to identify subtle structural features found in X-ray structures but not in native-like folds, since the latter will always have significant regions of local geometry atypical of real proteins. β-Strands, in particular, are not represented well by the four-state models used to generate the ensembles of plausible conformations (Park & Levitt, 1995). The diminished importance of distance dependence may be attributable to the same causes. The relative insignificance of the balance between hydrophobicity and residue-specificity in native-like identification is more difficult to understand. Perhaps native-like structures are simply too far from the correct structures for such effects to be seen.

When we turn to the ungapped threading test we find that the balance between distance dependence and hydrophobicity does not appear to be the determining factor. It is true that most of the VdW functions are better than the contact functions, and that the VdW(MJ)12 and VdW(HL)4 functions are the best of the VdW class, showing that these factors do seem to help. However, among the very best performers are the Shell and Shell(top) functions, neither of which are distance-dependent. The use of C^α centers in energy calculations seems to be counter-productive. Among the top performers, only the VdW(HL)4 function uses them. This observation is consistent with our explanation of the value of C^α centers for X-ray structure identification from the ensembles of plausible conformations. When segments of the sequence by chance are mounted upon their native secondary structure, many models are locally very similar to the correct structure, and using C^α atoms as an indicator for a native-like backbone is probably unnecessary. Furthermore, the superiority of the Shell function (a contact function without C^α centers) over the Contact(MJ) and Contact(HL) functions suggests that C^α centers are detrimental to a function's performance in the ungapped threading test. In summary, it appears that no function can simultaneously discriminate native structures from those that are native-like in their overall fold (i.e. the plausible ensemble set) and those that are native-like in local conformations (i.e. the ungapped threading set) because of the presence (or absence) of C^α interacting centers.

Understanding what function characteristic is important for successful discrimination of correct structures from the dynamics ensembles is much more straightforward. For the most part those functions that emphasize hydrophobic interactions have the best discriminating power. The function that is best at identifying the correct structure from the 298 K ensemble is the Surface function, followed fairly closely by HF(sm), VdW(MJ) and VdW(MJ)4. Several functions are extremely good discriminators for the 498 K ensembles. We still see that the best functions, in general, are those that have large hydrophobic components, such as the VdW(MJ), Surface, VdW(MJ)4, HF and HF(sm) functions. In the MD test, the HF(sm) improves upon the HF because its distance-dependent nature circumvents a problem associated with a "hard cutoff" in defining contacts: many of the near-

native conformations record favorable H-H contacts that are absent in the native structures (Huang *et al*., 1996). The weakest of all the functions include the VdW(HL) and VdW(HL)12 functions, which de-emphasize hydrophobicity the most of all the functions.

Why, then, does a strong hydrophobic component work so well for these dynamics ensembles? The MD simulations that generated these structures tended to expand the proteins relative to the starting X-ray structures. For instance, the radius of gyration for a 298 K structure increased by 2 to 3%, on average (data not shown). It is remarkable that simple functions such as the Surface and HF(sm) functions can detect these relatively small errors in the simulations.

Molecular expansion creates problems for other functions, however. Distance-dependent knowledge-based functions such as the Histogram function suffer from their use of statistics gathered from proteins of all sizes. Because the MD test set only comprises small proteins, the increase in interresidue distances associated with molecular expansion enhanced the scores of the decoys. When our parameter set for the Histogram, Shellm, and Shell(top)m functions were recompiled using only small proteins, the native fold rank scores improved for these functions, though not to the level of the hydrophobicity-emphasizing functions (data not shown). This phenomenon underscores the importance of proper compilation of parameter sets when using knowledge-based energy functions (Thomas & Dill, 1996).

Conclusions and future directions

This study has determined that functions that work well in one problem domain do not necessarily work well in others. Beyond this main conclusion, a number of other principles have been suggested. First, extraordinarily simple functions often work better than more complex functions: our results warn that investigators in structure prediction should be wary of additional complexity in their energy functions; it is often unnecessary and sometimes acts as a positive hindrance. Second, although the hydrophobic effect is certainly the largest force defining protein structure, it is not the only one: pairwise interactions do make a difference. Third, it is difficult for a single energy function to be effective at recognizing native-like features at both the local, secondary structure level and the global, tertiary structure level.

What are the practical implications for protein structure prediction? When attempting fold recognition (gapped threading), one should avoid the use of C^{α} interaction centers in addition to a virtual centroid coordinate. In the context of *ab initio* folding, where minimization of a target function produces a set of low-energy candidates, the data from the plausible ensembles suggest that there are several good choices, including the VdW(HL) and the Shell-type functions. Even though none of the functions can place all the near-native folds ahead of all the non-native folds on the score-sorted list, using one of the better functions as a filter will increase the effective concentration of near-native folds in the low-energy subset. Since some functions (e.g. VdW(HL)4 and VdW(MJ)12) place the native structure at an energy minimum with respect to the entire ensemble, it is possible, in principle, to couple these energy functions with a suitable search strategy to attain the native fold from the near-native folds in the subset obtained from the initial screen. The results from the molecular dynamics simulations suggest that functions that do not emphasize hydrophobicity should be used with caution in the late stages of a folding simulation since the final structures may not be sufficiently compact. Data from the MD test also highlighted another general principle: depending on the context, one should take protein size into account when compiling parameters for database-derived distance-dependent functions.

Methods

Database and test sets

Parameters for the various energy functions examined in this study were derived from the same database of well resolved X-ray structures described in our previous work (Park & Levitt, 1996). Ensembles of plausible secondary structure constrained protein conformations were generated for eight small proteins (Park & Levitt, 1996). Ungapped threading was performed on a set of 103 proteins described in earlier work (Huang *et al*., 1995). Two sets of 1000 conformations each for seven small proteins, were generated by molecular dynamics simulation at 298 K and 498 K (Huang *et al*., 1996). The functions examined in this study are summarized in Table 1 and the test sets outlined in Table 2. All protein structures and their identifiers were taken from the Brookhaven Protein Data Bank (PDB; Bernstein *et al*., 1977).

Energy functions

The Contact(MJ), Contact(HL), VdW(MJ), VdW(HL), Surface and Histogram energy functions have been described (Park & Levitt, 1996). The HF or hydrophobic fitness function has also been described in detail (Huang *et al*., 1995).

A major part of this study has been an examination of how changes in various energy functions effect discrimination for the different test sets. To that end, we introduce four new energy functions which are variants of the VdW energy functions, one which is a variant of the HF function, two which are empirically derived hydrophobic fitness functions, and four new contact potentials whose parameters are calculated differently from the previously described contact energy functions.

Contact(MJ), Contact(HL), VdW(MJ), VdW(HL) functions

The Contact(MJ) and Contact (HL) functions are on/off potentials of mean force derived from the frequencies of residue-residue contacts in a database of known protein structures. The Contact(MJ) function is derived similarly

to those by Miyazawa & Jernigan (1985, 1996). The open, unfolded state is the reference for this potential. The Contact(HL) potential is derived similarly to that by Hinds & Levitt (1992, 1994). The compact, randomly mixed state is the reference for this potential. Both the Contact potentials treat proteins as being constructed from 20 interacting centers, the 19 different side-chains (all residues except glycine) and a C^{α} backbone center. Both functions count contacts in the database of protein structures as those pairs of residues which have constituent atoms within 4 Å of each other.

The VdW(MJ) and VdW(HL) functions take the discrete interaction energies of the Contact(MJ) and Contact(HL) functions and fit them to a van der Waals-like functional form (see below).

Surface function

The Surface energy function is a crude estimate of the exposed hydrophobic surface area that treats each residue side-chain as a sphere whose volume is proportional to its atomic volume. C^{α} centers are also treated as simple spheres but are only used for determining the burial of side-chains (Park & Levitt, 1996). The exposed surface area of the side-chain is determined using the approximate method of Wodak & Janin (1980).

Histogram function

The Histogram energy function is a potential of mean force which, instead of using a single interaction energy for each residue pair type, assigns distinct energies to different separations in space and sequence (Park & Levitt, 1996; Sippl, 1990). This function's reference state is essentially the compact, randomly mixed state (but see Godzik *et al.*,1995).

VdW(MJ)12, VdW(MJ)4, VdW(HL)12 and VdW(HL)4 functions

The VdW(MJ) and VdW(HL) functions are derived from the Contact(MJ) and Contact(HL) functions. They have the form:

$$E = \sum_{i=1}^{N} \sum_{j=1+2}^{N} \frac{A_{ij}}{r_{ij}^{8}} - \frac{B_{ij}}{r_{ij}^{4}}$$

where N is the number of interacting centers in the protein, r_{ij} is the distance between interacting centers i and j, and A_{ij} and B_{ij} are parameters which depend on the types of centers i and j. The VdW(MJ)12 and VdW(HL)12 functions are also derived from the Contact(MJ) and Contact(HL) functions, respectively but have the functional form:

$$E = \sum_{i=1}^{N} \sum_{j=1+2}^{N} \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}$$

The VdW(MJ)4 and VdW(HL)4 functions are similarly derived with the functional form:

$$E = \sum_{i=1}^{N} \sum_{j=1+2}^{N} \frac{A_{ij}}{r_{ij}^{4}} - \frac{B_{ij}}{r_{ij}^{2}}$$

The VdW(MJ)12 and VdW(HL)12 functions have a sharper distance dependence than the original VdW functions, whereas the VdW(MJ)4 and VdW(HL)4 functions have a smoother distance dependence.

Shell and Shell(top) functions

All empirical energy potentials use one of two methods for deciding whether a pair of residues (or parts of residues) is in contact in a database of X-ray structures; this assignment is necessary for the statistical determination of energy parameters. The one used for the Contact and VdW functions regards a pair of residues as contacting if any one atom from one residue is within 4 Å of any one atom from the other residue. Hinds & Levitt (1992, 1994) used this technique also, using a 4.5 Å cutoff to derive their contact potential. For the VdW functions this methodology has the distinct advantage of allowing a simple method of estimating average contact distance between residue types (Park & Levitt, 1996). The other, and more common way, of determining residue contacts is by using a strict distance cutoff between a pair of side-chain centroids. In this study we generate four simple contact functions using this method. The functional form for the first of these potentials, which we designate the Shell energy function, is:

$$E = \sum_{i=1}^{N} \sum_{j=1+2}^{N} e_{ij}[\text{if } r_{ij} < 7.0 \text{ Å}]$$

where N is the number of residues in the protein, r_{ij} is the distance between the side-chain centroids of residue i and j, and the e_{ij} are contact energies derived using a 7 Å cutoff to define contacts in database proteins rather than atom-atom inter-residue contacts. The e_{ij} parameters are calculated as:

$$e_{ij} = -\ln(n_{ij}/n_{\exp ij})$$

where n_{ij} is the number of contacts closer than 7 Å between residues of types i and j found in the database, and n_{expij} is the expected number of contacts between residues of types i and j. The expected number of contacts of each type is calculated as:

$$n_{\exp ij} = \sum_{p} N_{pc} \frac{\sum_{k=1}^{N_p} \sum_{(l \geqslant 1) \cap (l \leqslant N_p) \cap ((l < k-1) \cup (l > k+1))} 1[\text{if } k \text{ is of type } i \text{ and } l \text{ is of type } j]}{(N_p - 2)(N_p - 1)}$$

where N_{pc} is the total number of contacts found in protein p and N_p is the number of residues in protein p. This expression calculates the expected number and kind of contacts, protein by protein, taking into account that nearest neighbors are excluded from contact. For all the Shell type functions the C^{α} backbone centers are not used.

The Shell(top) energy function is a variant of the Shell function in which the expected number of contacts for each residue pair type is calculated differently, namely as:

$$n_{\exp ij} = \sum_{p} N_{pc} \frac{\sum_{k=1}^{N_p} f_{|l-k|} \sum_{(l \geqslant 1) \cap (l \leqslant N_p) \cap ((l < k-1) \cup (l > k+1))} 1[\text{if } k = i \text{ and } l = j]}{(N_p - 2)(N_p - 1)}$$

where $f_{|l-k|}$ is the relative likelihood of contact between two residues separated by l-k amino acids in sequence. This relative likelihood is calculated from the database of

known protein structures by:

$$f_1 = \frac{n_l}{n_{\text{expl}}}$$

where n_l is the number of contacts between residues separated by l amino acids in sequence over the entire database, and n_{expl} is the number expected based on assuming a random distribution of contacts, which is calculated as:

$$n_{\text{expl}} = \frac{\sum_{j}^{N_p} \sum_{k=j+2}^{N_p} 1[\text{if } k-j=l]}{\frac{1}{2}(N_p-2)(N_p-1)}$$

The contact energies calculated using this method are at least partially corrected for the chain connectivity of polypeptide. Thomas & Dill (1996) showed that not taking chain connectivity into account results in systematic errors in energy parameters. The significance of these errors in discriminating correct from incorrect structures is unknown; our current results show little to distinguish the Shell and Shellm functions from the Shell(top) and Shell(top)m.

Of course, the normalization scheme presented here is not perfect. Ideally we would have liked to normalize expected contact frequencies for contacts between residue pairs of a given sequence separation using the unfolded state as a reference. This is impossible since the characteristics of this reference state are not known accurately enough. For this function and the Shell(top)m function described below, pseudo β-carbons located 3 Å from the α-carbon were used, instead of side-chain centroids. This choice makes the relative frequency of contacts between residues separated by different distances in sequence less noisy.

Shellm and Shell(top)m functions

The Shell and Shell(top) functions are not distance-dependent. For them, a pair of residues are either in contact or not. To address this possible deficiency we have devised versions of the Shell and Shell(top) energy functions which are, to some extent, distance-dependent. We designate them the Shellm and Shell(top)m functions. The functional form for both is:

$$\begin{aligned} E = &\sum_{i=1}^{N}\sum_{j=i+2}^{N} e_{6ij}[\text{if } r_{ij}<6\text{ Å}] \\ &+ \sum_{i=1}^{N}\sum_{j=i+2}^{N} e_{7ij}[\text{if } r_{ij}<7\text{ Å}] \\ &+ \sum_{i=1}^{N}\sum_{j=i+2}^{N} e_{8ij}[\text{if } r_{ij}<8\text{ Å}] \\ &+ \sum_{i=1}^{N}\sum_{j=i+2}^{N} e_{9ij}[\text{if } r_{ij}<9\text{ Å}] \\ &+ \sum_{i=1}^{N}\sum_{j=i+2}^{N} e_{10ij}[\text{if } r_{ij}<10\text{ Å}] \end{aligned}$$

This is simply the sum of five different Shell functions using a different distance cutoff for each. The individual e_{dij} parameters for the Shellm function are calculated the same way as for the Shell function, and the e_{dij} for the Shell(top)m function are calculated the same as for the Shell(top) function. The overlapping nature of the different contact energies which compose these functions tends to give higher weight to contacts that are near in space.

HF and HF(sm) functions

The hydrophobic fitness (HF) score is an exceedingly fast and simple energy function (Huang *et al*., 1995). Each amino acid residue is reduced to a virtual side-chain centroid 3.0 Å along the C^{α}-C^{β} vector and classified as either hydrophobic (H) or polar (P). The computation of the HF score is as follows:

$$HF = -\frac{\left(\sum_i B_i\right)\left(\sum_i (H_i - H_i^{\circ})\right)}{n^2}$$

where i is a hydrophobic (C, F, I, L, M, V, W) residue; B_i is a burial term, evaluated as the number of residues within 10 Å; H_i is the number of contacts made with non-polar side-chains (the seven hydrophobic residues plus Y), i.e. the number of non-polar centroids within 7.3 Å; n is the number of hydrophobic residues (excluding Y) in the sequence; and H_i^{o} is the number of hydrophobic contacts expected on a random basis.

In order to address some of the problems associated with on/off type potentials, we have modified the HF function to be distance-dependent. This continuous (or smooth) version of the HF score (designated HF(sm)) is a sigmoidal function weighted by a burial term:

$$HF_{\text{smooth}} = -\sum_i B_i \sum_{1\leqslant i\leqslant N} \sum_{i+2\leqslant j\leqslant N} \left(1/1+\frac{r_{ij}}{\sigma}\right)^n$$

where B_i is defined as above; i and j are hydrophobic residues (C, F, I, L, M, V, W); r_{ij} is the inter-residue distance (in Å); σ is the midpoint of the sigmoidal interaction curve, set at 7 Å; and n is a parameter that modulates the steepness of the curve, set at 10. Unlike Huang *et al*. (1996), the C^{β} positions for both these functions were computed as a functions of C^{α} atoms and not from the crystal structures. Minor differences in the data of Tables 5 to 7 and those found in Huang *et al*. (1995, 1996) reflect this procedural change, done for consistency with our other ensembles.

HF(stat) and HF(stat)m function

The HF(stat) energy function is a purely environmental function. That is, it does not depend on residue-residue interactions but only on the structural environment of individual residues. The form for this function is:

$$E = \sum_{i=1} e_{7i}(n_{7i})$$

where the e_{7i} are functions, specific for each amino acid type, of n_{7i}, the number of neighboring residues nearer than 7 Å. The e_{7i} parameters are derived from our database of known protein structures as:

$$\begin{aligned} e_{7i}(n) = &\\ &-\ln(\text{no. of times that residue type } i \\ &\text{has } n \text{ neighbors nearer than 7 Å}) \end{aligned}$$

For this function distances between residues are calcu-

lated between pseudo β-carbon atoms located 3 Å from the α-carbon positions.

The HF(stat)m function bears the same relationship to the HF(stat) functions as the Shellm and Shell(top)m functions. Its functional form is:

$$E = \sum_{i=1}^{N} e_{6i}(n_{6i}) + \sum_{i=1}^{N} e_{7i}(n_{7i}) + \sum_{i=1}^{N} e_{8i}(n_{8i}) + \sum_{i=1}^{N} e_{9i}(n_{9i}) + \sum_{i=1}^{N} e_{10i}(n_{10i})$$

Projection of energy function correlations to two dimensions

Although there are certain obvious correlations among the various energy functions we test in this paper, based either on their definitions, distance dependence, reference state, or performance, it is still desirable to have an objective method for determining how similar different energy functions are to each other. To this end we developed a simple strategy based on correlations in performance for pairs of energy functions over the different test sets.

We started by calculating the correlation coefficients between the average *Q*-scores (see below) for each test set over all pairs of energy functions. For this purpose we treated native-like identification and X-ray structure identification from our ensembles of plausible conformations as separate data points. We did the same for the 298 K and 498 K dynamics generated ensembles. Therefore there were five *Q*-scores for each energy function. What this procedure gave us was a matrix of correlation coefficients which each ranged from −1 to 1. What was needed was a matrix of pseudo distances between pairs of energy functions. Since correlation coefficients are in some sense analogous to dot products we transformed the matrix of coefficients to pseudo distances by taking the arccos of each of them. This yielded a matrix of distances all between 0 and π. We then took this matrix of pseudo distances and found the set of two-dimensional coordinates that best fits them in a least-squares sense. We used conjugate gradient minimization on the objective function:

$$F = \sum_{i=1}^{N} \sum_{j=i+1}^{N} (d_{ij} - r_{ij})^2$$

where d_{ij} is the pseudo distance between energy functions i and j, and r_{ij} is the distance between the two-dimensional coordinates of functions i and j. Minimizations were calculated from 200 sets of random starting coordinates. This is similar to the method use by Levitt (1983).

Measures of performance

We measure the quality of discrimination of energy functions with Rank scores, *Q*-scores, and *Z*-scores, all of which have been described (Park & Levitt, 1996; Huang *et al*., 1995, 1996). The Rank score is simply the energy rank of a target structure within a collection of other structures. The *Q*-score is a normalization of the Rank score which estimates the likelihood of obtaining a given rank relative to chance, taking into account the number of structures and the number of target structures within the collection. A *Q*-score increase of one corresponds to a tenfold improvement in discrimination. Placing a fold randomly in a score-sorted list corresponds to an average *Q*-score of 0.3. *Z*-scores are the extent to which energies depart from the mean energy (in standard deviation units) of a particular collection of structures. Thus, an average *Z*-score of zero is expected by chance.

Acknowledgements

This work was supported by the Department of Energy (grant number DE-FG03-95ER62135) and the National Institutes of Health (grant number GM 41455). We thank Jerry Tsai for running the molecular dynamics simulations described in this work.

References

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Jr, Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535–542.

Bowie, J. U. & Eisenberg, D. (1994). 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**, 4436–4440.

Bowie, J. U., Lüthy, R. & Eisenberg, D. (1991). A method to identify protein sequences that fold into a known three-dimensional structure. *Science,* **253**, 164–170.

Bryant, S. H. & Lawrence, C. E. (1993). An empirical energy function for threading protein sequence through folding motif. *Proteins Struct. Funct. Genet.* **16**, 92–112.

Cohen, F. E., Richmond, T. J. & Richards, F. M. (1979). Protein folding: evaluation of some simple rules for the assembly of helices into tertiary structures with myoglobin as an example. *J. Mol. Biol.* **132**, 275–288.

Covell, D. G. (1992). Folding protein α-carbon chains into compact forms by Monte Carlo methods. *Proteins: Struct. Funct. Genet.* **14**, 409–420.

Covell, D. G. (1994). Lattice model simulations of polypeptide chain folding. *J. Mol. Biol.* **235**, 1032–1043.

Dandekar, T. & Argos, P. (1994). Folding the main chain of small proteins with the genetic algorithm. *J. Mol. Biol.* **236**, 844–861.

Dandekar, T. & Argos, P. (1996). Identifying the tertiary fold of small proteins with different topologies from sequence and secondary structure using the genetic algorithm and extended criteria specific for strand regions. *J. Mol. Biol.* **256**, 645–660.

Fischer, D. & Eisenberg, D. (1996). Protein fold recognition using sequence-derived predictions. *Protein Sci.* **5**, 947–955.

Godzik, A., Kolinski, A. & Skolnick, J. (1992). Topology fingerprint approach to the inverse protein folding problem. *J. Mol. Biol.* **227**, 227–238.

Godzik, A, Kolinski, A. & Skolnick, J. (1995). Are proteins ideal mixtures of amino acids? Analysis of energy parameter sets. *Protein Sci.* **4**, 2107–2117.

Hendlich, M., Lackner, P., Weitckus, S., Floeckner, H., Froschauer, R., Gottsbacher, K., Casari, G. & Sippl, M. J. (1990). Identification of native protein folds amongst a large number of incorrect models. The calculation of low energy conformations from potentials of mean force. *J. Mol. Biol.* **216**, 167–180.

Hinds, D. A. & Levitt, M. (1992). A lattice model for protein structure prediction at low resolution. *Proc. Natl Acad. Sci. USA,* **89**, 2536–2540.

Hinds, D. A. & Levitt, M. (1994). Exploring conformational space with a simple lattice model for protein structure. *J. Mol. Biol.* **243**, 668–682.

Huang, E. S., Subbiah, S. & Levitt, M. (1995). Recognizing native folds by the arrangement of hydrophobic and polar residues. *J. Mol. Biol.* **252**, 709–720.

Huang, E. S., Subbiah, S., Tsai, J. & Levitt, M. (1996). Using a hydrophobic contact potential to evaluate native and near-native folds generated by molecular dynamics simulations. *J. Mol. Biol.* **257**, 716–725.

Jones, D. T., Taylor, W. R. & Thornton, J. M. (1992). A new approach to protein fold recognition. *Nature,* **358**, 86–89.

Kocher, J.-P.A., Rooman, M. J. & Wodak, S. J. (1994). Factors influencing the ability of knowledge-based potentials to identify native sequence-structure matches. *J. Mol. Biol.* **235**, 1598–1613.

Kolinski, A. & Skolnick, J. (1994). Monte Carlo simulations of protein folding. I. Lattice model and interaction scheme. *Proteins: Struct. Funct. Genet.* **18**, 338–352.

Levitt, M. (1983). Molecular dynamics of native protein. II. Analysis and nature of motion. *J. Mol. Biol.* **168**, 621–657.

Lathrop, R. H. & Smith, T. F. (1995). Global optimum protein threading with gapped alignment and empirical pair score functions. *J. Mol. Biol.* **255**, 641–665.

Maiorov, V. N. & Crippen, G. M. (1992). Contact potential that recognizes the correct folding of globular proteins. *J. Mol. Biol.* **227**, 876–888.

Miyazawa, S. & Jernigan, R. L. (1985). Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation. *Macromolecules,* **18**, 534–552.

Miyazawa, S. & Jernigan, R. L. (1996). Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *J. Mol. Biol.* **256**, 623–644.

Monge, A., Lathrop, E. J. P., Gunn, J. R., Shenkin, P. S. & Friesner, R. A. (1995). Computer modeling of protein folding: conformational and energy analysis of reduced and detailed protein models. *J. Mol. Biol.* **247**, 995–1012.

Mumenthaler, C. & Braun, W. (1995). Predicting the helix packing of globular proteins by sel-correcting distance geometry. *Protein Sci.* **4**, 863–871.

Needleman, S. B. & Wunsch, C. B. (1970). A general method applicable to the search for similarities in the amino acid sequences of two proteins. *J. Mol. Biol.* **48**, 443–453.

Ouzounis, C., Sander, C., Scharf, M. & Schneider, R. (1993). Prediction of protein structure by evaluation of sequence-structure fitness. Aligning sequences to contact profiles derived from three-dimensional structures. *J. Mol. Biol.* **232**, 805–825.

Park, B. H. & Levitt, M. (1995). The complexity and accuracy of discrete state models of protein structure. *J. Mol. Biol.* **249**, 493–507.

Park, B. & Levitt, M. (1996). Energy functions that discriminate X-ray and near-native folds from well-constructed decoys. *J. Mol. Biol.* **258**, 367–392.

Sippl, M. J. (1990). Calculation of conformational ensembles from potentials of mean force. An approach to the knowledge-based prediction of local structures in globular proteins. *J. Mol. Biol.* **213**, 859–883.

Sippl, M. J. (1995). Knowledge-based potentials for proteins. *Curr. Opin. Struct. Biol.* **5**, 229–235.

Sippl, M. J. & Weitckus, S. (1992). Detection of native-like models for amino acid sequences of unknown three-dimensional structure in a data base of known protein conformations. *Proteins: Struct. Funct. Genet.* **13**, 258–271.

Srinivasan, R. & Rose, G. D. (1995). LINUS: a hierarchic procedure to predict the fold of a protein. *Proteins: Struct. Funct. Genet.* **22**, 81–99.

Sun, S. (1993). Reduced representation model of protein structure prediction: statistical potential and genetic algorithms. *Protein Sci.* **2**, 762–785.

Sun, S., Thomas, P. D. & Dill, K. A. (1995). A simple protein folding algorithm using a binary code and secondary structure constraints. *Protein Eng.* **8**, 769–778.

Taylor, W. R. & Orengo, C. A. (1989). Protein structure alignment. *J. Mol. Biol.* **208**, 1–22.

Thomas, P. D. & Dill, K. A. (1996). Statistical potentials extracted from protein structures: how accurate are they? *J. Mol. Biol.* **257**, 457–469.

Vieth, M., Kolinski, A., Brooks, C. L., III & Skolnick, J. (1994). Prediction of the folding pathways and structure of the GCN4 leucine zipper. *J. Mol. Biol.* **237**, 361–367.

Vieth, M., Kolinski, A., Brooks, C. L., III & Skolnick, J. (1995). Prediction of quaternary structure of coiled coils. Applications to mutants of the GCN4 leucine zipper. *J. Mol. Biol.* **251**, 448–467.

Wallqvist, A. & Ullner, M. (1994). A simplified amino acid potential for use in structure prediction of proteins. *Proteins: Struct. Funct. Genet.* **18**, 267–280.

Wang, Y., Zhang, H., Li, W. & Scott, R. A. (1995a). Discriminating compact nonnative structures from the native structure of globular proteins. *Proc. Natl Acad. Sci. USA,* **92**, 709–713.

Wang, Y., Zhang, H. & Scott, R. A. (1995b). A new computational model for protein folding based on atomic solvation. *Protein Sci.* **4**, 1402–1411.

Wilmanns, M. & Eisenberg, D. (1993). Three-dimensional profiles from residue-pair preferences: identification of sequences with β/α-barrel fold. *Proc. Natl Acad. Sci. USA,* **90**, 1379–1383.

Wilson, C. & Doniach, S. (1989). A computer model to dynamically simulate protein folding: studies with crambin. *Proteins: Struct. Funct. Genet.* **6**, 193–209.

Wodak, S. J. & Janin, J. (1980). Analytical approximation to the accessible surface area of proteins. *Proc. Natl. Acad. Sci. USA,* **77**, 1736–1740.

Wodak, S. J. & Rooman, M. J. (1993). Generating and testing protein folds. *Curr. Opin. Struct. Biol.* **3**, 247–259.

Yue, K. & Dill, K. A. (1996). Folding proteins with a simple energy function and extensive conformational searching. *Protein Sci.* **5**, 254–261.

Edited by F. E. Cohen

(Received 23 August 1996; received in revised form 19 November 1996; accepted 25 November 1996)