A Test of Lattice Protein Folding Algorithms

K Yue, KM Fiebig, PD Thomas, HS Chan, EI Shakhnovich, and KA Dill

PNAS 1995;92;325-329
doi:10.1073/pnas.92.1.325

Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 325–329, January 1995
Chemistry

A test of lattice protein folding algorithms

KAIZHI YUE*, KLAUS M. FIEBIG*, PAUL D. THOMAS†, HUE SUN CHAN*, EUGENE I. SHAKHNOVICH‡, AND KEN A. DILL*

*Department of Pharmaceutical Chemistry, Box 1204, and †Graduate Group in Biophysics, Box 0448, University of California, San Francisco, CA 94143-1204; and ‡Department of Chemistry, Harvard University, 12 Oxford Street, Cambridge, MA 02138

Communicated by Peter G. Wolynes, University of Illinois, Urbana, IL, August 31, 1994 (received for review February 7, 1994)

ABSTRACT **We report a blind test of lattice-model-based search strategies for finding global minima of model protein chains. One of us (E.I.S.) selected 10 compact conformations of 48-mer chains on the three-dimensional cubic lattice and used their inverse folding algorithm to design HP (H, hydrophobic; P, polar) sequences that should fold to those "target" structures. The sequences, but not the structures, were sent to the UCSF group (K.Y., K.M.F., P.D.T., H.S.C., and K.A.D.), who used two methods to attempt to find the globally optimal conformations: "hydrophobic zippers" and a constraint-based hydrophobic core construction (CHCC) method. The CHCC method found global minima in all cases, and the hydrophobic zippers method found global minima in some cases, in minutes to hours on workstations. In 9 out of 10 sequences, the CHCC method found lower energy conformations than the 48-mers were designed to fold to. Thus the search strategies succeed for the HP model but the design strategy does not. For every sequence the global energy minimum was found to have multiple degeneracy with 10^3 to 10^6 conformations. We discuss the implications of these results for (*i*) searching conformational spaces of simple models of proteins and (*ii*) how these simple models relate to proteins.**

Computer algorithms are emerging that attempt to predict the three-dimensional structures of proteins from their amino acid sequences (1–7). The best blind test of a folding algorithm is the prediction of a protein structure that is already known to someone but is not known to the predictor. Recently, proteins have been modeled at low resolution as chains configured on spatial lattices. Algorithms have arisen for inverse folding (8, 9) that design sequences to fold to a desired given conformation and for folding (1, 3, 10–12) that take sequences and predict their native states. The virtue of lattice models is that their native states can often be known exactly, many of their properties are well understood, and in many respects, they resemble those of real proteins. The best consistency check of lattice model folding algorithms would be if someone "inverse folded" a protein (i.e., designed a sequence to fold to a known native state) and gave it to a "folder" to attempt to predict its native state.

This is the idea behind the present paper. This work began as a friendly wager. E.I.S. (representing the Harvard group) proposed to design some 48-mer HP sequences (H, hydrophobic; P, polar; see refs. 13–15) that would fold to three-dimensional simple cubic lattice target structures of his choice and to give the sequences to the UCSF group (K.Y., K.M.F., P.D.T., H.S.C., and K.A.D.). The UCSF group would then attempt to fold the sequences to the best possible structures, based on the HP potential. By fold, we refer here only to thermodynamics and not to kinetics: we mean that an algorithm finds the lowest energy state, without consideration of whether there is kinetic access to that state. For the present study, if structures with energies equal to or better than the energy of the target (designed) structure were found, then this would be considered successful folding. The UCSF group would use two strategies: hydrophobic zippers (HZ) (10) and a systematic assembly process using discrete geometry (constraint-based hydrophobic core construction, CHCC) (12, 16). The Harvard group offered a six pack of beer if the UCSF group could successfully fold 1 or more out of 10 sequences of 48-mers. This is a report of the results.

METHODS

HP Lattice Model. Proteins used in the HP lattice model are specific sequences of H (hydrophobic) and P (polar) monomers (13, 15, 17) configured on the three-dimensional simple cubic lattice. Each chain configuration is a self-avoiding walk on the lattice. Contacts between H monomers are favorable. The energy E of a chain conformation is determined by the number of H–H contacts h: $E = -|\varepsilon|h$, where $|\varepsilon|$ is a positive constant. (In this work E is given in units of $|\varepsilon|$.) The native state of an HP sequence is defined as the set of conformation(s) with the largest possible number h of H–H contacts. The energy of a native conformation is defined as E_N, and the number of conformations with this energy is the native state degeneracy g_N.

Sequence Design. Inverse folding was performed by the Harvard group by iterative Monte Carlo interchanging of monomers, for a given target structure, until convergence to a low-energy sequence was achieved. We call these target structures "putative native states" (PNS) because initially it was not known whether the designed sequences could fold to conformations with lower energies than the PNS energy. The details of the design method are given in refs. 9 and 18. Ten sequences designed this way were sent to the UCSF group. These sequences are listed in Fig. 1. The UCSF group was given only the HP monomer sequences and not the target structures, so that it could be a legitimate blind test. All 10 designed structures chosen by the Harvard group are maximally compact (see Fig. 2). The Harvard group also gave the UCSF group the PNS energy E_{PNS} so that the UCSF group could know if it had succeeded in meeting the criterion of reaching either the PNS or a structure at least as good energetically.

Folding of Designed Sequences. The UCSF group used two procedures to find native states for the given sequences. The first, HZ (10, 11), is an opportunistic process that begins with randomly chosen H–H contacts that can be formed among near neighbors in the sequence and zips up other H–H contacts as they come into spatial proximity by virtue of preceding contacts.

Two properties of HZ have previously been found. (*i*) HZ can find global minima of short HP lattice model chains for some sequences without exhaustive searching of conforma-

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: HP, hydrophobic polar; HZ, hydrophobic zippers; CHCC, constraint-based hydrophobic core construction; PNS, putative native states.

	Sequence (PNS Conformation)	E_{PNS}
1.	HPHHPPHHHHPHHHPPHHPPHPHHHPHPHHPPHHPPPHPPPPPPPPHH	-30
	(RFFRBULBULDFFFRBUFLBBRRFRBBLDDRUFDFULFURDDLLLBB)	
2.	HHHHPHHPHHHHHPPHPPHHPPHPPPPPPHPPHPPPHPPHHPPHHHPH	-30
	(RFFLBUFRBRDBRFFFLBUULBLFFDDRUURDRUBDBULBRDLLULD)	
3.	PHPHHPHHHHHHPPHPHPPHPHHPHPPPHPPHHPPHHPPHPHPPHP	-31
	(RFUBUFFRBDFDRUUBBUFFLLLDDRDLBUBUFUBRFRBDDRFDBLF)	
4.	PHPHHPPHPHHHPPHPHHPPPHHHHHPPHPHHPHPHPPPPHPPHPHP	-30
	(RRFLLUFDRUUFRBBBDLLURFDRFDFRBUBDBUUFFFDLLDLUUBB)	
5.	PPHPPPHPHHHHPPHHHHPHHPHHHPPHPHPHPPHPPPPPPHHPHHPH	-30
	(RFLFUBBRFFFLDRBRUULBRDDRBLUULLFFFRRRBBBDFFFLDRB)	
6.	HHHPPPHHPHPHHPHHPHHPHPPPPPPPHPHPPHPPPHPPHHHHHHPH	-30
	(RFFRBUUULFDBBDLUFFUBBRRRFFLDRBBLDDRUFDFULLBLFDB)	
7.	PHPPPPHPHHHPHPHHHHPHHPHHPPPHPHPPPHHHPPHHPPHHPPPH	-31
	(RRRFULDLUULFURDDLDBUBRULUFRBRRDDLUFRULFRDLDRDLL)	
8.	PHHPHHHPHHHHPPHHHPPPPPPHPHHPPHHPHPPPHHPHPHPHHPPP	-31
	(RRRFLUUFDRDFULLBLBBUFFRBDDLFRRFLLUURRRBBDBLLURR)	
9.	PHPHPPPPHPHPHPPHPHHHHHHPPHHHPHPPHPHHPPHPHHHPPPPH	-30
	(RRFFRUULDLLUUBRFDBRDBRDFUUBUFFLBBDLULDDRFDFLBUU)	
10.	PHHPPPPPPHHPPPHHHPHPPHPHHPPHPPHPHHPPHHHHHHHPPHH	-30
	(RFLFUUUBRBLDFRRBRFFUBBLFFLDRDLDRBRFUBBDLUFLLBRU)	

FIG. 1. HP sequences and PNS conformations of the 10 Harvard sequences. The conformations are encoded in bond directions (U, up; D, down; L, left; R, right; F, forward; B, backward). A three-dimensional rendering of PNS conformations of sequences 2, 8, and 10 is shown in Fig. 2.

tional space (10). (*ii*) For longer chains, this method explores compact conformations that have much hydrophobic clustering but does not always find native states, perhaps because of computational limitations (19). Thus the present test is a useful challenge for the method.

The second UCSF folding method, CHCC, is based on the geometric properties of lattice chains (16). This approach consists of (*i*) an optimization process that first determines a tight upper bound on the number h of possible H–H contacts (i.e., lower bound for the energy of the native state E_N) and then (*ii*) construction of such conformations by assembling a core of H residues with a minimal surface area, subject to several constraints imposed by the sequence. This method determines rigorously when conformations are at their global minima. In addition to constructing native conformations CHCC also gives a lower bound on the native state degeneracy g_N. The method is described in ref. 16, and the present implementation is in ref. 12.

RESULTS

For each of the 10 designed sequences, Table 1 lists the PNS energy E_{PNS}, the lowest energy found by the HZ method E_{HZ}, and the native state energy E_N and the lower bound on native degeneracy g_N obtained by the CHCC approach. In Fig. 2 for sequences 2, 8, and 10, we show the PNS conformation and a representative native conformation found by the HZ and CHCC algorithms. We found the following results.

The HZ Method Quickly Finds Conformations with the PNS Energy. We generated 10,000 HZ pathways for each of the 10 sequences. HZ succeeded in 7 cases out of 10 in finding conformations at least as good as the PNS conformations to which they were designed to fold. For the remaining three sequences, Monte Carlo kinetics of the Harvard group (18, 20) found conformations having lower energy than any of the 10,000 HZ conformations (Table 2).

There Are Many Different Conformations That Have the PNS Energy. HZ performed at UCSF and Monte Carlo kinetics performed at Harvard turned up hundreds of conformations of each sequence that had the PNS energy. Thus the design strategy is not creating sequences that will uniquely fold to the target structures.

The reason for the design problems could be any of the following possibilities, each of which is considered in more detail in the *Discussion* below. (*i*) The simple interaction in the HP model is too nonspecific for most sequences to have unique native structures. Thus it would be impossible to design a sequence, in the HP model, by any procedure, that would have a unique native structure. In this connection, it is known that using a multiletter code, rather than the binary HP code used here, or even allowing both favorable and unfavorable intermonomer interactions fosters native uniqueness and diminishes the degeneracy of low-energy conformations (refs. 18 and 21 and H.S.C., unpublished results). (*ii*) The sequence design algorithm used here is not optimal. It may not be sufficient to merely interchange monomers during the design process, keeping the HP composition fixed. (*iii*) It may be that not every possible target conformation is "designable" in the HP model. Earlier studies (8, 15) found that some compact conformations simply cannot be uniquely encoded in any HP sequence. If so, target conformations cannot be chosen arbitrarily, and sequence design also requires concurrent structure design in the HP model.

The Target Structures Are Not Global Energy Minimum Conformations of the Sequences. For 9 of the 10 sequences, the CHCC approach (16) found conformations lower in energy than the target conformations to which they were designed to fold ($E_{PNS} > E_N$). HZ found lower energies in 4 out of 10 cases. The Monte Carlo procedure found none of lower energy. The true global minimum energies E_N (determined by the CHCC method) are given in Table 1, and some representative conformations are shown in Fig. 2 along side the target conformations. Thus the designed sequences will not fold to their target structures because they are not stable states of those sequences. In Table 1, we also show the CHCC-determined lower bounds on the numbers of different conformations at the

Table 1. Comparison of PNS energy E_{PNS}, the lowest energy obtained by 10,000 HZ runs E_{HZ}, and the native state energy E_N and the lower bound g_{CHCC} on native degeneracy g_N by CHCC

Seq.	E_{PNS}	E_{HZ}	E_N	g_{CHCC}
1	−30	−31	−32	1.5×10^6
2	−30	−32	−34	14×10^3
3	−31	−31	−34	5×10^3
4	−30	−30	−33	62×10^3
5	−30	−30	−32	54×10^3
6	−30	−29	−32	52×10^3
7	−31	−29	−32	59×10^3
8	−31	−29	−31	306×10^3
9	−30	−31	−34	10^3
10	−30	−33	−33	188×10^3

Table 2. HZ end-state distributions for the 10 sequences at various energies

		Number of HZ end states					
Seq.	E_{HZ}	E_{HZ}	E_{HZ} + 1	E_{HZ} + 2	E_{HZ} + 3	E_{HZ} + 4	E_{HZ} + 5
1	−31	1	8	84	423	1248	2134
2	−32	1	5	99	425	1145	2030
3	−31	6	44	191	689	1752	2674
4	−30	2	23	210	706	1667	3010
5	−30	2	30	166	548	1263	2093
6	−29	4	27	157	674	1788	2845
7	−29	12	55	239	782	1525	2395
8	−29	8	81	352	1060	2108	3186
9	−31	21	126	510	1272	2219	2587
10	−33	1	3	55	175	528	1073

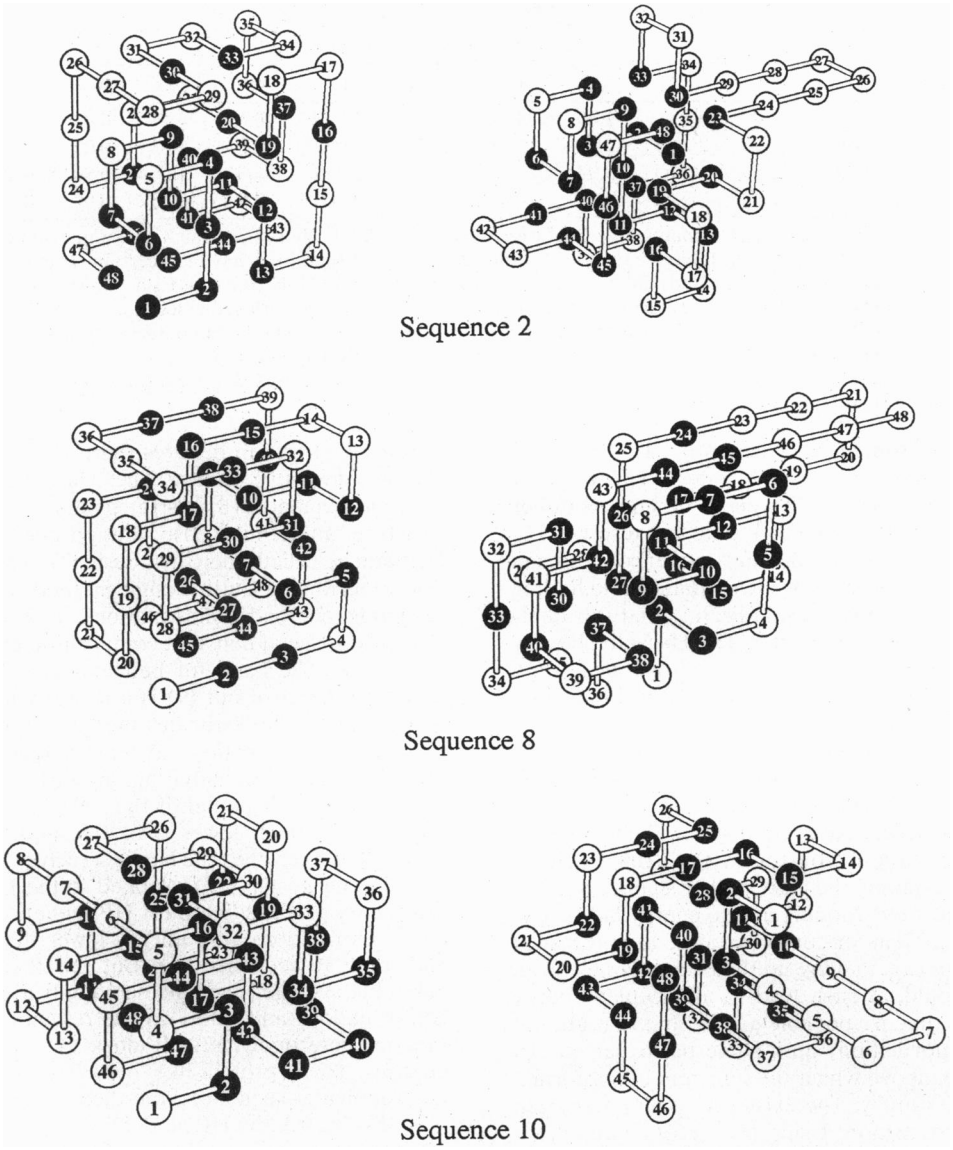

FIG. 2. For sequences 2, 8, and 10, the respective PNS (*Left*) conformations and selected native conformations (*Right*) are shown.

global energy minimum (native state degeneracies g_N). There are at least approximately 10^3 to 10^6 global minima for each of these HP sequences.

DISCUSSION

What do we learn from these results? First, the number of global minima for these 10 48-mer HP sequences in three dimensions is larger than 10^3. Random pairwise structural comparisons indicate that two ground-state conformations of the same sequence have only 35–55% of H–H contacts in common, on average (Table 3). This implies that these 10 sequences do not fold uniquely as biological sequences do. But biological uniqueness seems likely to require some design, and appropriate designs may not be encoded within these 10 sequences. What does "unique" mean? At higher resolution, the "unique" native states of real proteins are themselves ensembles involving small fluctuations around native-like energies and structures (22, 23). We believe that in coarse-grained lattice models, these fluctuations should be captured largely within a single lattice conformation. Hence, the conclusion that these 10 sequences do not fold uniquely refers to large-scale structural diversity of the ground states, not to small perturbations around a single "fold." We point out that other theorists are more agnostic about the nature of fluctuations in native proteins (24). Honeycutt and Thirumalai (25) have argued in agreement with Frauenfelder *et al.* (26) that the tier-zero substates or so-called taxonomic substates of folded proteins may well correspond to slightly different overall folds.

How should we regard this high degeneracy of native structures of these 10 sequences? Is ground-state conformational diversity characteristic of real random-sequence polypeptides? We do not know. Designed polypeptides often do not fold to unique structures (27). For both real polypeptides and lattice model sequences, there are good designs, with little native conformational diversity, and bad designs (8). We have found some HP sequences of 60–80 monomers that have fewer than five native conformations (12), although we have not yet found any that have only a single native conformation. These particular HP sequences mimic real protein sequences in having very limited native conformational diversity. Replacing two-letter codes (H and P, for example) by multiletter codes undoubtedly helps reduce degeneracy, and this may be more protein-like.

Second, the Harvard sequence design procedure does not work for HP lattice model chains. The Harvard and UCSF

Table 3. Comparison of ground-state conformations

	% pairs with given structural similarity										
Seq.	100%	90%	80%	70%	60%	50%	40%	30%	20%	10%	0%
1	0.26	0.27	0.67	1.69	5.22	18.52	38.49	29.63	5.26	0.00	0.00
3	1.86	1.78	1.73	2.84	11.97	22.76	41.87	13.32	1.85	0.01	0.00
7	7.00	12.86	10.65	9.94	17.36	12.78	11.39	16.37	1.65	0.00	0.00
10	1.89	0.94	3.52	4.16	9.24	38.60	33.68	7.93	0.03	0.00	0.00

For each sequence, 399 ground state conformations are randomly chosen for pairwise comparison of H–H contact maps. Each table entry indicates the percentage of such pairs having a certain structural similarity (0–100% as indicated), in terms of percentages of common H–H contacts. Each column of the table indicates the range of similarity in increments of 10%. For example, for sequence 3, 2.84% of pairs have 70–80% of their H–H contacts that are identical. Note that the same H–H contact map does not usually map to the same conformation and does not even necessarily map to the same shape of the H core. This is because the H residues are usually not connected among themselves. So, when the H–H contact maps of two conformations differ, the conformations differ considerably.

groups interpret this somewhat differently. The Harvard group believes that this same design strategy will work for more protein-like models having 20 types of monomers rather than 2 (18) and that it does not work here because the HP model is not sufficiently protein-like. For sequences designed by this method using a 20-letter code (18), Monte Carlo folding simulations always lead to the PNS, indicating that it is likely to be the true global energy minimum. The Harvard strategy is based solely on the target structure and regards a mean-field argument, described below, as insurance that no other conformations will be populated.

Real proteins may have larger "energy gaps" than the HP model (9, 18). To describe energy gaps, consider an energy level diagram. The set of native conformation(s) has the lowest energy and thus they define the ground state. The set of next lowest-energy conformations constitutes the "first excited" state, the conformations of the next higher energy are the "second excited" state, etc. Moving up the energy ladder can lead to increasing numbers of conformations. In contrast, if there were an energy gap, moving up the energy ladder from the ground state would, at first, lead to unoccupied energy levels (i.e., energies that are possible *a priori* but that are not adopted by any conformation) until some particular energy level, E_C, is reached, above which the numbers of conformations would increase rapidly. The existence of an energy gap implies a two-state transition. Using the CHCC method, we find that none of the 10 sequences has an energy gap; i.e., the "first-excited" $E = E_N + 1$ levels of all 10 sequences are occupied. However, it is the full density of states that fundamentally determines stability. In some applications, the latter should be studied rather than just considering "gap" or "no gap."

For real proteins, which have local "excitations" (fluctuations) near the global energy minimum (22, 23), the notion of energy gap becomes more complex than in lattice models with discrete energy levels. In some spin-glass model of proteins, the energy gap is defined to be the difference between the lowest energy of the random energy landscape and the energy of the minimally frustrated native state (28–30). Energy gaps and other spin-glass concepts have also been applied to protein structure prediction in conjunction with associate-memory Hamiltonians and with local Hamiltonians similar to those used here (31, 32). We do not know how prevalent or rare energy gaps might be in three-dimensional HP lattice proteins, just as we do not know how common they are in random polypeptide sequences. In two dimensions, short-chain exhaustive simulations show that the fraction of sequences that have unique ground states (i.e., a single lowest-energy lattice conformation) is approximately 2.5% (15, 33), and so far we have found only six 18-monomer HP sequences in two dimensions with both unique ground states and energy gaps (33). Much less is known about three dimensions.

The UCSF group believes there may be a problem in the Harvard design procedure. There are two criteria for designing good sequences (8): a strategy must "design in" the target structure, and it must also "design out" bad conformations (sometimes called "negative design"). The UCSF group believes the Monte Carlo design method has no provision for designing out bad conformations: it is an optimization in sequence space at constant composition but not in conformation space. The basis for believing that bad conformations would be designed out (9) is a mean-field theory of heteropolymers (34) that estimates the lower bound on energies of nonnative conformations of an "average" sequence with a given interaction potential among different monomer types. According to this model, if the PNS energy of the designed sequence is much lower than this bound, then the target structure is likely the true native structure: unique, stable, and with an energy gap (18). Applied to the HP interaction (18), this theory predicts that most HP sequences will have numerous conformations with energy lower than the PNS energy, and these 10 sequences do. But this theory applies only to ensemble-averaged sequences and only to maximally compact conformations and relies on the resemblance of proteins to random heteropolymers. It does not apply to specific sequences, so if protein-like behavior is a rare feature of heteropolymer sequences, the theory would not apply.

In this regard, the HP model is at one end of a spectrum of models: it has considerable conformational diversity at energies close to the global minimum, and yet, as indicated here, it is possible to devise algorithms to find some global minima conformations based only on sequence information. In models with multiletter codes and energy gaps, perhaps simpler Monte Carlo strategies (18) can design and fold model proteins, since the landscapes are less rugged. The CHCC and HZ methods are currently applicable only to the HP model.

This collaboration was restricted to the question of whether particular computer algorithms, namely, HZ and CHCC, could find global minima of energy landscapes in HP lattice models. We did not consider other models of intrachain interactions, other definitions of native states of proteins, or kinetic issues of how folding might be limited by kinetic traps on the conformational space, although it is clear that sequences with large energy gaps (in several models) may be both fast folding and stable at certain optimal intrachain interactions (9, 18, 20, 33). A stronger hypothesis has been suggested: that the existence of an energy gap is a sufficient condition for a sequence to fold rapidly (9, 18, 20).

CONCLUSIONS

While this challenge between Harvard and UCSF groups was originally undertaken in fun, we believe there are several useful conclusions that come from it. (*i*) The 10 HP lattice sequences in this study have multiply-degenerate ground states

and do not have unique native structures as in real proteins, but they may represent the kind of conformational diversity expected in real polypeptides of random amino acid sequences. Although this study does not rule out the existence of HP sequences that have unique three-dimensional native structures with energy gaps, it does suggest they are probably rare. It is not known whether these properties are rare or common among real polypeptides. Several lattice and spin-glass models, including this one, indicate that the ability to fold to unique states is not a property of arbitrary copolymer sequences; it requires some design, biological or otherwise. An interesting unsolved question is, What is the minimal number of monomer types that would allow for unique, stable, and accessible native conformations for at least a biologically significant fraction of sequences and conformations? (*ii*) The present study shows that arbitrary HP sequences can be folded to many of their global minima in minutes on present computer workstations. (*iii*) Our results imply that some care is required to design sequences that do not fold to incorrect conformations. Real proteins fold to unique stable accessible native states. Good search strategies should be able to find such states for models of rugged energy landscapes.

The UCSF group thanks the National Institutes of Health and the Office of Naval Research for financial support. P.D.T. is a Howard Hughes Medical Institute Predoctoral Fellow. E.I.S. acknowledges numerous fruitful discussions with Alexander Gutin and Peter Leopold and financial support from the Packard Foundation. Sci. USA **90,** 1942–1946.

1. Kolinski, A. & Skolnick, J. (1993) *J. Chem. Phys.* **98,** 7420–7433.
2. Piela, L., Kostrowski, J. & Scheraga, H. (1989) *J. Phys. Chem.* **93,** 3339–3346.
3. Ma, J., Hsu, D. & Straub, J. (1993) *J. Phys. Chem.* **99,** 4024–4035.
4. Cohen, F. E., Sternberg, M. J. E., Phillips, D. C., Kuntz, I. D. & Kollman, P. A. (1980) *Nature (London)* **286,** 632–634.
5. Sun, S. (1993) *Protein Sci.* **2,** 762–785.
6. Argos, P. (1987) *J. Mol. Biol.* **197,** 331–348.
7. Finkelstein, A. V. & Reva, B. (1991) *Nature (London)* **351,** 497–499.
8. Yue, K. & Dill, K. A. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 4163–4167.
9. Shakhnovich, E. I. & Gutin, A. M. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 7195–7199.
10. Fiebig, K. M. & Dill, K. A. (1993) *J. Chem. Phys.* **94,** 3475–3487.
11. Dill, K. A., Fiebig, K. M. & Chan, H. S. (1993) *Proc. Natl. Acad.*
12. Yue, K. & Dill, K. A. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 146–150.
13. Lau, K. F. & Dill, K. A. (1989) *Macromolecules* **22,** 3986–3997.
14. Lau, K. F. & Dill, K. A. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 638–642.
15. Chan, H. S. & Dill, K. A. (1991) *J. Chem. Phys.* **95,** 3775–3787.
16. Yue, K. & Dill, K. A. (1993) *Phys. Rev. E* **48,** 2267–2278.
17. Shortle, D., Chan, H. S. & Dill, K. A. (1992) *Protein Sci.* **1,** 201–215.
18. Shakhnovich, E. I. (1994) *Phys. Rev. Lett.* **72,** 3907–3910.
19. Lattman, E. E., Fiebig, K. M. & Dill, K. A. (1994) *Biochemistry* **33,** 6158–6166.
20. Sali, A., Shakhnovich, E. I. & Karplus, M. (1994) *J. Mol. Biol.* **235,** 1614–1636.
21. O'Toole, E. M. & Panagiotopoulos, A. Z. (1992) *J. Chem. Phys.* **97,** 8644–8652.
22. Frauenfelder, H., Sligar, S. G. & Wolynes, P. G. (1991) *Science* **254,** 1598–1603.
23. Straub, J. E. & Thirumalai, D. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 809–813.
24. Wolynes, P. G. (1992) in *Spin Glass and Biology*, ed. Stein, D. (World Scientific, Singapore), pp. 225–259.
25. Honeycutt, J. D. & Thirumalai, D. (1992) *Biopolymers* **32,** 695–709.
26. Frauenfelder, H., Alberding, N. A., Ansari, A., Braunstein, D., Cowen, B. R., *et al.* (1990) *J. Phys. Chem.* **94,** 1024–1037.
27. Handel, T. M., Williams, S. A. & DeGrado, W. F. (1993) *Science* **261,** 879–885.
28. Bryngelson, J. D. & Wolynes, P. G. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 7524–7528.
29. Bryngelson, J. D. & Wolynes, P. G. (1989) *J. Phys. Chem.* **93,** 6902–6915.
30. Bryngelson, J. D. & Wolynes, P. G. (1990) *Biopolymers* **30,** 177–188.
31. Goldstein, R. A., Luthey-Schulten, Z. A. & Wolynes, P. G. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 4918–4922.
32. Goldstein, R. A., Luthey-Schulten, Z. A. & Wolynes, P. G. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 9029–9023.
33. Chan, H. S. & Dill, K. A. (1994) *J. Chem. Phys.* **100,** 9238–9257.
34. Shakhnovich, E. I. & Gutin, A. M. (1989) *Biophys. Chem.* **34,** 187–199.