Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 2536–2540, April 1992
Biophysics

A lattice model for protein structure prediction at low resolution

(protein folding/tertiary structure/conformational search)

D. A. HINDS AND M. LEVITT

Beckman Laboratories for Structural Biology, Department of Cell Biology, Stanford University School of Medicine, Stanford, CA 94305

Communicated by Aaron Klug, December 23, 1991

ABSTRACT **The prediction of the folded structure of a protein from its sequence has proven to be a very difficult computational problem. We have developed an exceptionally simple representation of a polypeptide chain, with which we can enumerate all possible backbone conformations of small proteins. A protein is represented by a self-avoiding path of connected vertices on a tetrahedral lattice, with several amino acid residues assigned to each lattice vertex. For five small structurally dissimilar proteins, we find that we can separate native-like structures from the vast majority of non-native folds by using only simple structural and energetic criteria. This method demonstrates significant generality and predictive power without requiring foreknowledge of any native structural details.**

The three-dimensional structures of protein molecules are thought to be largely if not completely determined by their amino acid sequences (1, 2). However, the prediction of structure from sequence has proved to be extremely difficult. Protein structure prediction must overcome two related problems: (*i*) the size of a protein's conformational space and (*ii*) the presence of local minima in a protein's potential energy landscape. The number of possible conformations accessible to even a small protein is so vast that an exhaustive conformational search for global energy minima will never be possible. In turn, exploration of any small region of conformational space will identify local energy minima that slow down directed strategies that attempt to move systematically toward the most stable structure.

For any prediction strategy, there is a trade-off between the accuracy of the protein representation and the amount of conformational space that can be searched. A protein in the course of folding explores only a tiny fraction of conformational space, because the folding process must be guided by the same sorts of intramolecular interactions that stabilize the final folded structure (3, 4). Molecular dynamics with detailed interatomic potential functions can realistically model proteins on the picosecond to nanosecond time scale (5). It is doubtful that such detailed simulations will ever be able to model folding, which occurs on a time scale of milliseconds or seconds (6–8). Similarly, energy-minimization strategies that employ detailed structural models are capable of identifying global energy minima only for very small systems (9, 10). Simplified representations of polypeptide chains with fewer degrees of freedom can be used to improve the sampling of conformational space (11, 12). Some of the most restrictive models constrain the paths of polypeptides to pass through points on a lattice (13–16). Despite their simplifications, these models have not permitted exhaustive conformational searches and so far have not been able to progress from a realistic protein sequence toward a correct structure in the absence of any assumed structural information.

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.

The conformation of a properly folded protein is stabilized by the coordination of many specific atomic interactions. The cost of using a simplified model of protein structure is the loss of some fraction of this detailed structural information. It is not clear how simple a model of a protein can be and still retain enough of this information to allow some discrimination of good from bad folds.

We have developed an exceptionally simple lattice model of protein structure for which we can exhaustively evaluate all possible folds for small proteins. This model avoids the conformational complexity problem by capturing only the large-scale features of a protein fold and avoids the problem of local energy minima because it is feasible to estimate the conformational energies of all possible structures. We find that for a variety of small proteins we can reliably separate correctly folded structures from the vast majority of misfolded conformations. Our model has predictive power and at the same time does not require the knowledge of any structural information about a protein beyond its amino acid sequence.

METHODS

A Lattice Model for Protein Structures. We have chosen to represent a polypeptide chain as a self-avoiding chain of connected vertices on a bounded diamond-like tetrahedral lattice. A unique feature of our model is that we do not enforce a one-to-one correspondence between lattice vertices and residues. One lattice vertex can represent several residues, and a model structure contains half as many vertices as there are residues in the sequence. This representation bears little resemblance to real protein structures and cannot accurately represent α-helices or β-strands. However, we have found that a path on this lattice is sufficiently flexible to capture the range of possible topologies of a polypeptide backbone. In fact, fitting every other residue to a vertex seems to be a better match to the overall flexibility of polypeptides than fitting every residue.

We represent a small protein of 60 amino acid residues by a self-avoiding path of 30 lattice vertices. For this small system, there are still on the order of $3^{29} \approx 10^{13}$ possible lattice structures. To reduce this to a manageable number, we require that our structures be reasonably compact and globular. Each structure is required to fit within a predefined eliptical bounding volume at least 50% larger than the volume occupied by the protein, and to have a radius of gyration close to that of a sphere with equal volume. We also eliminate degenerate symmetry-related structures. With the restrictions we typically use, only around 10^7 lattice structures jointly satisfy these constraints, and it is computationally possible to exhaustively search all of them.

Conformational Energy Calculations. Statistical estimates of the effective interaction energies for all residue pairs are calculated from the observed frequencies of contacts in x-ray structures (Fig. 1). Our data base included 56 structures

Abbreviations: drms, root-mean-square distance deviation; BPTI, bovine pancreatic trypsin inhibitor.

totaling more than 12,000 residues and specifically excluded the five proteins under study. The method is similar to, but simpler than, that used by Miyazawa and Jernigan (17). We define x-ray contacts wherever a heavy atom of one residue comes within 4.5 Å of a heavy atom belonging to another residue. Because this lattice model cannot accurately represent the exposed surface of a structure, we do not explicitly include the effects of solvent interactions. The energy (e_{uv}) of a contact between residue types u and v is estimated from the number of residue–residue contacts (N_{uv}), the total number of residues (N_p), and the number of residues of each type (N_u, N_v) in each separate protein:

$$e_{uv} = -\log\left(\frac{1}{\sum_p N_p}\sum_p \frac{N_p N_{uv}}{N_u N_v}\right), \quad [1]$$

where p varies over all proteins in the data set. We change the effective interaction energy for a cysteine–cysteine contact to that of a cysteine–serine contact, to better reflect only noncovalent interactions. The energy (C_{ij}) of a contact between two vertices i and j, when they are mapped to sequence positions m_i and m_j, is calculated using

$$C_{ij} = \frac{2e_{r_{m_i}r_{m_j}} + e_{r_{m_i-1}r_{m_j}} + e_{r_{m_i+1}r_{m_j}} + e_{r_{m_i}r_{m_j-1}} + e_{r_{m_i}r_{m_j+1}}}{6}, \quad [2]$$

where r_{m_i} is the residue type of residue m_i in the sequence. This gives some weight to interactions with flanking residues that do not map directly onto vertices, but not as much as to the central interaction. In energy calculations, interactions are defined between all pairs of occupied vertices that are either nearest (4.95 Å) or next-nearest (8.08 Å) neighbors.

An assignment of exactly two residues to each vertex is not necessarily optimal, and we have devised a method to align an amino acid sequence to a particular chain path. We first choose a default assignment of every second residue to a vertex (Fig. 2). We calculate the contribution to the conformational energy of each vertex with its current mapping, and what its contribution would be if all else stayed the same but the mapping was shifted to one side or the other. We use a dynamic programming strategy to find the combination of mutually compatible moves that together yield the best total conformational energy. Mappings are restricted so that one residue cannot occupy multiple vertices, and a maximum of three residues can be squeezed into the gap between two vertices. The ends of the chain are also free to shift, but at most three residues can extend past a terminal vertex. This corresponds roughly to the range of inter-residue spacings that can be achieved in real proteins between a fully extended chain and an α-helix. The shifts are applied to the current mapping, and the process is repeated until either there are no more good moves or a maximum number of iterations is reached. The result of the procedure is a list of residue numbers $m_1 \ldots m_n$ for a walk with n steps.

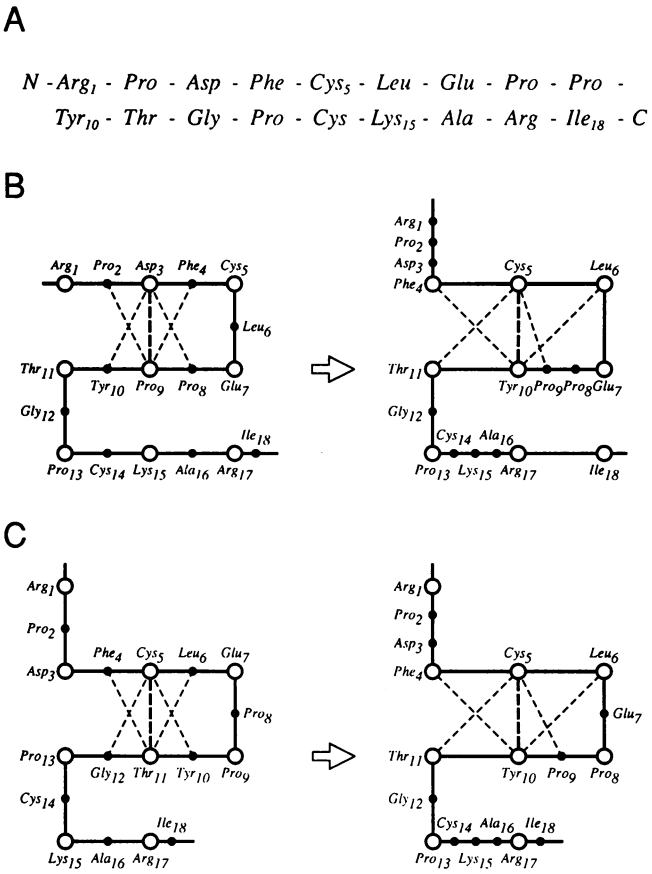

FIG. 2. Illustration of alignment optimization process in two dimensions. The first 18 residues of bovine pancreatic trypsin inhibitor (BPTI) (*A*) are mapped to a simple nine-step walk (*B Left*), so that initially every other residue is assigned to a vertex. The residue–residue interactions contributing to one lattice contact energy are shown with dashed lines. After four cycles of sliding the sequence along the walk to optimize the energy, the alignment stabilizes (*B Right*). The most obvious change is that many hydrophobic residues have shifted onto vertices. If the alignment process is applied to a different walk with a similar overall shape (*C Left*), the sequence converges to a virtually identical mapping (*C Right*). The alignment finds the same set of inter-residue contacts despite the fact that the two mappings are initially completely different.

We use rms distance deviations (drms values) to compare lattice walks with known structures. The drms for a path of length n passing through vertices $w_1 \ldots w_n$, for a particular residue mapping $m_1 \ldots m_n$, is calculated using

	ALA	CYS	ASP	GLU	PHE	GLY	HIS	ILE	LYS	LEU	MET	ASN	PRO	GLN	ARG	SER	THR	VAL	TRP	TYR
ALA	0.18	-0.01	0.12	0.62	-0.76	0.40	-0.47	-0.92	0.28	-0.61	-0.87	0.28	0.34	0.31	0.08	0.06	-0.09	-0.63	-0.76	-0.77
CYS	-0.01	-0.27	0.26	-0.15	-1.45	0.12	-1.37	-0.56	0.11	-0.67	-1.52	0.12	0.45	-0.77	-1.40	-0.27	0.12	-0.75	-1.48	-0.90
ASP	0.12	0.26	0.86	0.16	0.11	0.40	-1.37	0.39	-0.09	0.92	0.14	0.24	1.17	0.29	-0.89	0.19	-0.25	0.35	-0.27	-0.29
GLU	0.62	-0.15	0.16	-0.57	-0.30	1.06	-0.76	0.57	-0.84	0.42	0.08	0.00	0.50	-0.12	-0.76	0.23	-0.08	0.57	-0.65	-0.56
PHE	-0.76	-1.45	0.11	-0.30	-1.65	-0.03	-1.27	-1.60	0.05	-1.43	-1.52	-0.22	0.27	-0.31	-0.81	-0.37	-0.90	-1.23	-1.46	-0.89
GLY	0.40	0.12	0.40	1.06	-0.03	0.00	0.15	0.01	0.63	0.09	-0.35	0.62	0.36	0.26	-0.18	0.44	0.40	0.06	-0.70	0.24
HIS	-0.47	-1.37	-1.37	-0.76	-1.27	0.15	-0.95	-0.04	0.72	0.29	-1.20	-0.49	-0.28	-0.38	-0.42	-0.26	0.11	0.24	-1.38	-0.83
ILE	-0.92	-0.56	0.39	0.57	-1.60	0.01	-0.04	-1.49	0.20	-1.49	-1.60	0.15	0.44	0.21	0.05	0.15	-0.39	-1.34	-1.48	-0.85
LYS	0.28	0.11	-0.09	-0.84	0.05	0.63	0.72	0.20	1.13	0.28	0.48	0.54	0.91	0.31	1.18	0.61	0.15	0.77	-0.24	-0.57
LEU	-0.61	-0.67	0.92	0.42	-1.43	0.09	0.29	-1.49	0.28	-1.33	-1.11	0.26	0.46	0.19	-0.28	0.15	-0.13	-1.09	-1.07	-0.71
MET	-0.87	-1.52	0.14	0.08	-1.52	-0.35	-1.20	-1.60	0.48	-1.11	-1.51	-0.81	-0.04	0.72	-0.77	0.08	-0.63	-1.10	-1.91	-0.88
ASN	0.28	0.12	0.24	0.00	-0.22	0.62	-0.49	0.15	0.54	0.26	-0.81	-0.25	0.56	-0.30	0.06	-0.20	0.18	0.23	-0.49	-0.59
PRO	0.34	0.45	1.17	0.50	0.27	0.36	-0.28	0.44	0.91	0.46	-0.04	0.56	0.77	-0.51	-0.25	0.48	0.63	0.07	-0.58	-0.53
GLN	0.31	-0.77	0.29	-0.12	-0.31	0.26	-0.38	0.21	0.31	0.19	0.72	-0.30	-0.51	0.20	-0.30	0.80	0.00	0.00	0.05	-1.04
ARG	0.08	-1.40	-0.89	-0.76	-0.81	-0.18	-0.42	0.05	1.18	-0.28	-0.77	0.06	-0.25	-0.30	-0.64	-0.26	0.48	0.08	-1.00	-1.01
SER	0.06	-0.27	0.19	0.23	-0.37	0.44	-0.26	0.15	0.61	0.15	0.08	-0.20	0.48	0.80	-0.26	-0.06	-0.05	0.11	-0.23	-0.28
THR	-0.09	0.12	-0.25	-0.08	-0.90	0.40	0.11	-0.39	0.15	-0.13	-0.63	0.18	0.63	0.00	0.48	-0.05	-0.26	-0.31	-0.10	-0.36
VAL	-0.63	-0.75	0.35	0.57	-1.23	0.06	0.24	-1.34	0.77	-1.09	-1.10	0.23	0.07	0.00	0.08	0.11	-0.31	-1.26	-1.13	-0.67
TRP	-0.76	-1.48	-0.27	-0.65	-1.46	-0.70	-1.38	-1.48	-0.24	-1.07	-1.91	-0.49	-0.58	0.05	-1.00	-0.23	-0.10	-1.13	-1.04	-0.63
TYR	-0.77	-0.90	-0.29	-0.56	-0.89	0.24	-0.83	-0.85	-0.57	-0.71	-0.88	-0.59	-0.53	-1.04	-1.01	-0.28	-0.36	-0.67	-0.63	-0.40

FIG. 1. Matrix of estimated pairwise interaction energies in *RT* units.

Table 1. Attributes of proteins used for evaluating the prediction scheme

	3RXN	4PTI	1R69	1SN3	1CTF
Residues	52	58	63	65	68
Structural class	All-β	$\alpha+\beta$	All-α	$\alpha+\beta$	α/β
Constraints	4*	3†	None	4†	None
Walk length	26	29	31	33	34
Fixed mapping					
Average drms, Å	5.65	6.99	5.43	6.64	6.48
Lowest drms, Å	3.24	3.93	3.17	4.01	3.90
With optimization					
Average drms, Å	5.64	6.49	5.28	6.34	6.02
Lowest drms, Å	3.07	3.56	2.87	3.67	3.28

The reference structures are identified by their Protein Data Bank designations. The proteins are rubredoxin (3RXN), BPTI (4PTI), the N-terminal domain of 434 repressor (1R69), scorpion neurotoxin variant 3 (1SN3), and the C-terminal domain of ribosomal protein L7/L12 (1CTF). Structural classes were assigned as in ref. 19. The lower part of the table illustrates how our alignment optimization process improves the rms deviations between the populations of bounded walks and native structures, compared to using a fixed mapping of every second residue to a lattice vertex.
*Cys—Fe bonds in iron binding site.
†Cys—Cys disulfide bonds.

$$\text{drms} = \sqrt{\frac{\sum_i \sum_{j \neq i} (D_{w_i w_j} - D_{m_i m_j})^2}{n(n-1)}}, \quad [3]$$

where $D_{w_i w_j}$ is the distance between the vertices w_i and w_j, and $D_{m_i m_j}$ is the C^α distance between residues m_i and m_j in the reference structure. Only the subset of residues mapped to vertices contribute to this deviation. While rms coordinate deviations are generally preferred (18), distance deviations are better at measuring the progress of our method. Our prediction method cannot distinguish between a structure and its mirror image, and the drms shares this symmetry.

RESULTS

We chose five small proteins of known structure, spanning all major structural classes (19), to evaluate our model (Table 1). All coordinates were taken from the Protein Data Bank (20). These proteins range from 52 to 68 residues, corresponding to lattice structures of from 26 to 34 steps. We used a bounded lattice containing 50 vertices to model all these proteins (Fig. 3). A lattice edge length of 4.95 Å gave the best scaling between lattice coordinates and native C^α positions. This is only 0.875 times the value expected from amino acid volume data (21, 22), which would be the best scaling if the C^α positions of real proteins were evenly distributed throughout their interiors. As larger proteins are studied, the best scaling would be expected to move closer to the ideal value.

Evaluation of the Lattice Representation. The choice of a particular bounding shape may arbitrarily restrict the range of shapes that lattice paths can assume. We have tried to verify that our 50 vertex bounded lattice is free of bias by using BPTI as a test case. We searched the walk populations of several other bounded tetrahedral lattices with from 43 to 54 vertices and with different elliptical shapes. Although different native-like walks were found for each lattice, the fraction of native-like structures was nearly independent of the bounding shape in this range. The 50-vertex lattice was able to represent all the proteins in our test set with similar levels of fidelity (Table 1). In each case, the best structures have drms values at least 5 standard deviations better than the averages over all bounded lattice structures.

Our strategy of dynamically optimizing the mapping of sequences to lattice structures significantly improves the rms fit of the most native-like structures (Table 1). While our strategy generally identifies only locally optimal mappings, we have found that stronger optimization methods do not significantly improve the results over those of our simple approach. This optimization does not require any native structural information—it depends only on a protein's amino acid sequence and the contact energy parameters. Because this procedure effects this improvement by varying the spacing of residues along the chain path, it must be extracting a kind of secondary structural information from the sequence despite the fact that interactions between residues close together along the sequence are explicitly neglected.

Prediction of Native Folds. We tested the feasibility of using our model for structure prediction by seeing how far we could go toward predicting the native structure of BPTI by stepwise application of structural and energetic filters (Table 2). We empirically chose the stringencies of each type of restriction to yield the greatest net enrichment of native-like structures, judged by their drms values relative to the native structure. We found that we could reduce the population of 10^7 bounded walks to <500, with a 7000-fold increase in the proportion of walks meeting our criterion for native-likeness. The best lattice structures included in the predicted walk group are quite successful in capturing the overall chain fold of BPTI (Fig. 4), though some details are lost.

As an unbiased way of assessing the generality of our prediction scheme, we applied the same selection parameters we had determined for BPTI to the other proteins in our test set. We increased the stringency of the radius-of-gyration cutoff for proteins larger than BPTI, because larger proteins typically show less irregularity of shape than the smallest ones. We also increased the stringency of the energy-per-

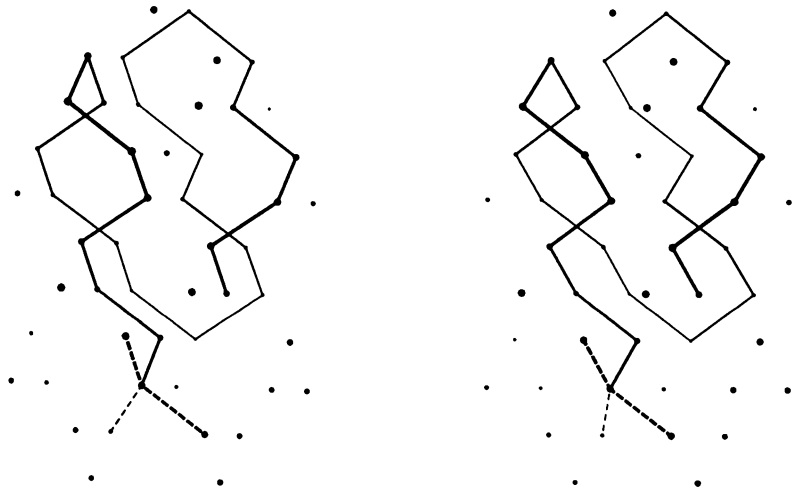

FIG. 3. A step in assembling a lattice structure. The solid dots represent vertices of a bounded tetrahedral lattice. The lattice was constructed by enumerating all vertices contained in an ellipsoid with major axes equal to 8, 8, and 10.8 lattice edge lengths. This set of 50 vertices was used to generate all the lattice structures described in this paper. As a path is traced through the lattice, there are at most 3 lattice positions available for extending the path at each step that are not excluded by the self-avoidance requirement. The dashed lines represent the possible next steps for this path.

Table 2. "Purification table" for BPTI

Selection step*	No. of walks		Enrichment of native folds	
	Total	Native†	This step	Total
Unique bounded‡	11,429,748	30		
Optimization§	11,429,748	970	32×	32×
Radius of gyration¶	2,018,558	588	3.4×	111×
Disulfide bonds‖	63,918	236	13×	1407×
Energy per contact**	420	8	5.2×	7257×

*The selections were applied sequentially in the order shown.
†Walks with drms < 4.2 Å from the native structure.
‡Symmetry-unrelated walks that fit within the bounded 50-vertex lattice.
§Sequence alignment adjusted by dynamic programming to minimize conformational energy.
¶Radius of gyration less than 1.12 times that of a sphere with the same volume.
‖Walks for which the average distance between cysteine pairs that are crosslinked in the native structure was less than 2.0 lattice spacings.
**Walks whose energies were at least 2 standard deviations below the average for walks meeting the radius-of-gyration criterion.

contact selection for the proteins that had no disulfide constraints. We did not adjust any structural or energetic parameters of the model. The results for all the test proteins were similar to those for BPTI (Fig. 5 and Table 3). The two energy-driven steps—the alignment optimization and the energy-per-contact selection—account for most of the predictive power of the method. The success of the method also shows no consistent dependence on problem size.

Perhaps our most surprising finding is that our primitive conformational energy function has significant discriminating power, despite the low resolution of the lattice model. Because our model does not explicitly account for solvent and because the numbers of walk contacts are so variable, the average energy per contact is the most robust parameter for classifying structures. For each of the test proteins, this parameter is strongly correlated with a walk's similarity to the native structure. This is especially remarkable, given that of the contacts contributing to a walk's conformational energy, rarely more than a third represent genuine native interactions.

DISCUSSION

One advantage of predicting structure via an exhaustive search is the simplicity of measuring the progress of the prediction scheme. We can characterize a protein's entire conformational space and measure the effectiveness of each

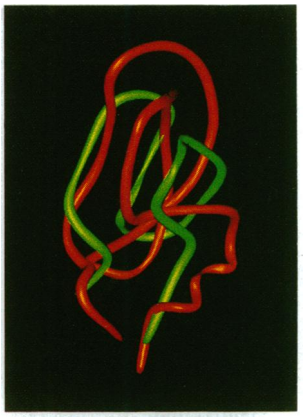

FIG. 4. Native and lattice structures of BPTI. The native (red) and lattice (green) structures are represented by ribbons passing through their C^{α} positions. The best lattice structures of BPTI are quite successful in capturing the protein's overall fold. This particular lattice structure has a rms coordinate deviation of 5.65 Å, and a drms of 3.89 Å. Of the interactions contributing to the lattice structure's conformational energy, 22% are genuine native contacts. This structure was contained in the set of 420 conformations that satisfied all the structure-prediction criteria.

predictive step as a sort of "purification" of correct, native folds. It is doubtful that our structure prediction strategy could reliably identify a single best structure for a particular sequence. It seems clear, however, that it is sufficiently powerful to eliminate the vast majority of candidate folds. The enrichment is efficient enough that a second stage of analysis could afford to spend several thousand times as much computational effort per structure in further narrowing the search. Our best structures compare very favorably to those found by other groups using more faithful models of protein structure. The best of our 420 BPTI structures has a drms deviation of 3.9 Å, whereas previous studies have reported values of 5.3 Å (11), 4.7 Å (23), and 4.5 Å (24).

An important feature of our model is that conformations can be generated and evaluated extremely quickly. Using a Silicon Graphics Iris 4D-240 workstation, we can analyze 10 million BPTI structures in 80 min. Though the choice of a particular bounding volume directly limits the numbers of structures to be searched, the computational demands of our method will ultimately show an exponential dependence on problem size, because larger bounding shapes will be necessary to model larger proteins. The simplicity of our model is crucial here, because the exponential term scales with lattice path length rather than the larger sequence length. We expect

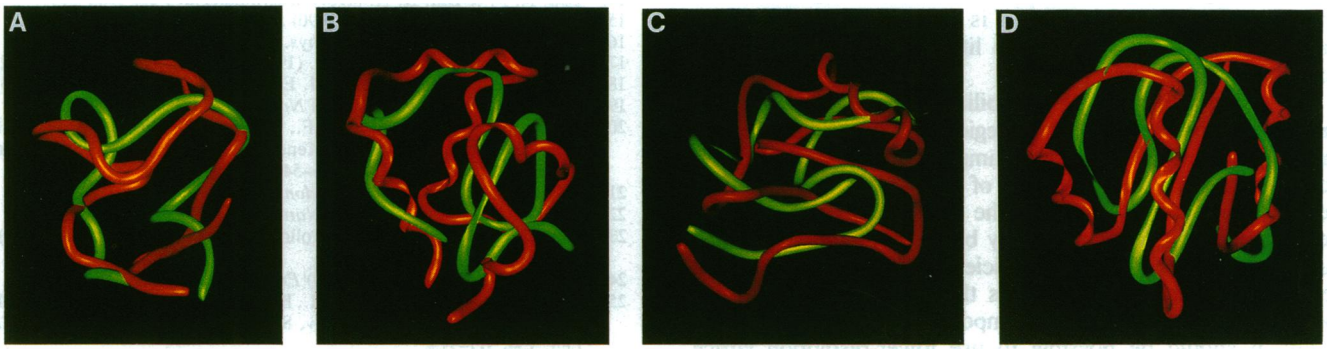

FIG. 5. Comparison of native and lattice structures for each test protein. C^{α} ribbons from the native x-ray structures of rubredoxin (*A*), λ repressor (*B*), scorpion neurotoxin (*C*), and the C-terminal fragment of ribosomal protein L7/L12 (*D*) are shown in red, and selected native-like lattice structures are in green. The four lattice structures have drms values of 3.67, 3.19, 3.93, and 3.97 Å relative to their corresponding native C^{α} positions. Each of these structures has a low conformational energy and is contained in the set of structures that satisfy all the selection criteria of the structure-prediction method. As in the case of BPTI, the lattice structures are capable of capturing the overall fold of each protein, with some distortion and loss of detail.

Table 3. Results of applying BPTI selection criteria to other proteins

	3RXN	4PTI	1R69	1SN3	1CTF
drms cutoff, Å	3.4	4.2	3.3	4.2	4.0
Before selection					
Total walks	7.1×10^6	1.1×10^7	1.3×10^7	1.2×10^7	1.1×10^7
Native-like	6	30	32	14	16
Selection step					
Optimization	19×	32×	21×	64×	133×
Radius of gyration	2.2×	3.4×	4.4×	2.5×	2.5×
Disulfide bonds	4.4×	13×	n/a	6.0×	n/a
Energy per contact	20×	5.2×	20×	14×	5.7×
Total enrichment	3649×	7267×	1849×	13,082×	1911×
After selection					
Total walks	3228	420	872	676	1410
Native-like	10	8	4	10	4

The drms cutoffs used to track the purification of native-like structures were chosen to make the numbers of native-like structures in each unprocessed walk population similar. The radius-of-gyration limit was lowered to 1.10 times that of a sphere with equal volume for 1SN3, 1R69, and 1CTF. Because 1R69 and 1CTF lack disulfide constraints, we increased the strength of their energy per contact selection and retained walks at least 3 standard deviations better than average. In every case, the original population of roughly 10^7 walks was reduced to on the order of 10^4 structures, of which a much larger fraction were native-like.

to be able to evaluate structures as large as 80–100 residues, comparable to single domains of larger proteins.

While our model can capture the general fold and arrangement of intramolecular contacts of a structure, it contains no information about specific atomic interactions. The early stages of folding may be guided by relatively nonspecific and time-averaged interactions of the sort that can be represented by this model. The range of conformations that can be built on our lattice could be compared to the range of distinguishable structures of a molten globule (25–27), in which a compact, weakly ordered structure has formed but specific tertiary contacts have not yet stabilized.

A general prediction strategy must not depend on the foreknowledge of any specific structural information about the proteins whose structures are to be predicted. We have been very careful to avoid including any such requirements in our method. We allow the use of disulfide constraints because these linkages can often be determined experimentally without actually solving a protein's three-dimensional structure (28, 29). When using more accurate structural models, other groups have generally been forced to make simplifying assumptions by using extra information derived from the same structures they then try to predict. This information has included assumptions of secondary structure (15), or assumptions of precise native molecular shapes (13), or selection of structural and energetic parameters for a specific target structure (11). It is interesting how little specific structural information is necessary to recreate a native fold, but this may have little bearing on a general solution to the folding problem.

We are investigating the possibility of integrating other secondary-structure prediction strategies with the alignment optimization procedure. A simple example of this idea would be to add cooperativity to the choice of how many residues to fit between two lattice vertices. The next step toward a more detailed structure prediction may be to subdivide our lattice, such that the vertices are sufficiently dense that they can represent side chains as well as the polypeptide backbone. While such a lattice would be impossible to traverse exhaustively, it should be possible to use lower-resolution lattice structures as templates to generate small families of similar, but more detailed, forms. A more sophisticated potential function could then distinguish between side chain–side chain, side chain–backbone, and backbone–backbone interactions. A lattice capable of representing the orientations of side chains might also be sufficiently realistic to justify including solvent interactions. Given the success of this method at low resolution, the possibility of similar results at higher resolution with more faithful lattice models is very encouraging.

We thank S. Subbiah and V. Daggett for help with this manuscript. D.A.H. is a Howard Hughes Medical Institute Predoctoral Fellow. This work was supported by the National Institutes of Health, Grant GM41455 (M.L.). Color graphics images were generated using the MidasPlus software system from the Computer Graphics Laboratory, University of California, San Francisco.

1. Anfinsen, C. B., Haber, E., Sela, M. & White, F. H. (1961) *Proc. Natl. Acad. Sci. USA* **47**, 1309–1314.
2. Anfinsen, C. B. (1973) *Science* **181**, 223–230.
3. Levinthal, C. (1968) *J. Chim. Phys.* **65**, 44–45.
4. Wetlaufer, D. B. (1973) *Proc. Natl. Acad. Sci. USA* **70**, 697–701.
5. Levitt, M. & Sharon, R. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 7557–7561.
6. Baldwin, R. L. (1975) *Annu. Rev. Biochem.* **44**, 453–475.
7. Udgaonkar, J. B. & Baldwin, R. L. (1988) *Nature (London)* **335**, 694–699.
8. Roder, H., Elöve, G. A. & Englander, S. W. (1988) *Nature (London)* **335**, 700–704.
9. Scheraga, H. A. & Paine, G. H. (1986) *Ann. N.Y. Acad. Sci.* **482**, 60–68.
10. Li, Z. & Scheraga, H. A. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 6611–6615.
11. Levitt, M. & Warshel, A. (1975) *Nature (London)* **253**, 694–698.
12. Levitt, M. (1976) *J. Mol. Biol.* **104**, 59–107.
13. Covell, D. G. & Jernigan, R. L. (1990) *Biochemistry* **29**, 3287–3294.
14. Lau, K. F. & Dill, K. A. (1989) *Macromolecules* **22**, 3986–3997.
15. Skolnick, J. & Kolinski, A. (1990) *Science* **250**, 1121–1125.
16. Gō, N. (1983) *Annu. Rev. Biophys. Bioeng.* **12**, 183–210.
17. Miyazawa, S. & Jernigan, R. L. (1985) *Macromolecules* **18**, 534–552.
18. Cohen, F. E. & Sternberg, M. J. E. (1980) *J. Mol. Biol.* **138**, 321–333.
19. Levitt, M. & Chothia, C. (1976) *Nature (London)* **261**, 552–558.
20. Bernstein, F. C., Koetzle, T. F., Williams, G. J. D., Meyer, E. F., Brice, M. D., Rodgers, G. R., Kennard, O., Shimanochi, T. & Tasumi, M. (1977) *J. Mol. Biol.* **112**, 535–542.
21. Chothia, C. (1975) *Nature (London)* **254**, 304–308.
22. Chothia, C. & Janin, J. (1975) *Nature (London)* **256**, 705–708.
23. Kuntz, I. D., Crippen, G. M., Kollman, P. A. & Kimelman, D. (1976) *J. Mol. Biol.* **106**, 983–994.
24. Wilson, C. & Doniach, S. (1989) *Proteins* **6**, 193–209.
25. Dolgikh, D. A., Gilmanshin, R. I., Brazhnikov, E. V., Bychkova, V. E., Semisotnov, G. V., Venyaminov, S. Yu. & Ptitsyn, O. B. (1984) *FEBS Lett.* **136**, 311–315.
26. Ohgushi, M. & Wada, A. (1983) *FEBS Lett.* **164**, 21–24.
27. Kuwajima, K. (1989) *Proteins* **6**, 87–103.
28. Brown, J. R. & Hartley, B. S. (1966) *Biochem. J.* **101**, 214–228.
29. Carlsson, S. R. & Fukuda, M. (1989) *J. Biol. Chem.* **264**, 20526–20531.