Building native protein conformation from highly approximate backbone torsion angles

Haipeng Gong, Patrick J. Fleming, and George D. Rose*

T. C. Jenkins Department of Biophysics, The Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218

Communicated by S. Walter Englander, University of Pennsylvania School of Medicine, Philadelphia, PA, September 26, 2005 (received for review August 2, 2005)

Reconstructing a protein in three dimensions from its backbone torsion angles is an ongoing challenge because minor inaccuracies in these angles produce major errors in the structure. As a familiar example, a small change in an elbow angle causes a large displacement at the end of your arm, the longer the arm, the larger the displacement. Even accurate knowledge of the backbone torsions ϕ and ψ is insufficient, owing to the small, but cumulative, deviations from ideality in backbone planarity, which, if ignored, also lead to major errors in the structure. Against this background, we conducted a computational experiment to assess whether protein conformation can be determined from highly approximate backbone torsion angles, the kind of information that is now obtained readily from NMR. Specifically, backbone torsion angles were taken from proteins of known structure and mapped into 60° × 60° grid squares, called mesostates. Side-chain atoms beyond the β-carbon were discarded. A mesostate representation of the protein backbone was then used to extract likely candidates from a fragment library of mesostate pentamers, followed by Monte Carlo-based fragment-assembly simulations to identify stable conformations compatible with the given mesostate sequence. Only three simple energy terms were used to gauge stability: molecular compaction, soft-sphere repulsion, and hydrogen bonding. For the six representative proteins described here, stable conformers can be partitioned into a remarkably small number of topologically distinct clusters. Among these, the native topology is found with high frequency and can be identified as the cluster with the most favorable energy.

protein structure | protein secondary structure | protein fragment assembly | Monte Carlo simulation

Protein molecules are known to undergo a reversible disorder ⇄ order transition (1). In the classical view, the unfolded state is thought to be a structurally featureless ensemble that adopts its native structure spontaneously and uniquely under conditions that favor folding. For many small proteins of biophysical interest, this well studied folding reaction is apparently a two-state process. Accordingly, it can be represented by the equation: U(nfolded) ⇄ N(ative), with equilibrium constant, $K_{\mathrm{eq}} = \mathrm{N}/\mathrm{U}$, and free energy, $\Delta G^0 = -\mathrm{RT}\ln K_{\mathrm{eq}}$, the free energy difference between the two populations. Central to this view, U is thought to be largely comprised of randomly coiled molecules, and N is thought to be largely comprised of uniquely structured molecules.

Lately, we have been exploring the doubly divergent alternative view that the unfolded state is more organized (2) and the folded state is less homogeneous (3) than previously thought. If so, then both states can be treated productively as constrained thermodynamic ensembles. Numerous recent papers suggest that the unfolded population is not featureless, despite the fact that it does indeed exhibit random-coil statistics (4). Based on residual dipolar couplings from NMR, Shortle (5) and Shortle and Ackerman (6) have argued that the denatured state retains native-like topology, although not all agree (7–9). Another hypothesis regards unfolded proteins as fluctuating ensembles of polyproline II helix (10–18).

Our recent work indicates that steric clash (19) and hydrogen bonding (20, 21) promote organization in proteins, effectively eliminating many conceivable random-coil conformers (22, 23). These two organizing factors influence folded and unfolded states alike. Consistent with this conclusion, the coil library (24), a subset of the folded population, is hypothesized to represent the unfolded population (25). More often than not, a pronounced bias toward native-state secondary structure can be detected in the local amino acid sequence by Monte Carlo simulations that emphasize sterics and hydrogen bonding (26). Even random-coil statistics (27), long taken to be the hallmark of the unfolded state (28), do not preclude the possibility that the unfolded state is far from featureless (29). All of these lines of evidence converge on the idea that the unfolded state is more organized than previously thought.

Here, we shift our focus to the native state and attempt to assess heterogeneity in the folded population. Starting with a highly approximate backbone conformation, such as might be available from NMR or calculated biases, we use a Monte Carlo algorithm to produce compact chain conformations subject to three simple filters involving (*i*) global compaction, (*ii*) steric exclusion, and (*iii*) hydrogen bonding. Conceptually, our algorithm seeks to maximize backbone hydrogen bonding, subject to the constraint that the resultant structure is compact, but not unrealistically so. Successful folds elaborated in this way are then clustered by structural criteria, not by energy, and the clusters are enumerated and classified. Our approach is deliberately reminiscent of early work of Richards and coworkers (30).

Surprisingly, only a small number of clusters is probable under these simple constraints. Usually, though not invariably, the largest cluster corresponds to the native state. The algorithm operates solely on the backbone, where it is presumed that conformational biases are largely exerted via local side-chain:backbone interactions (ref. 26 but see also ref. 31). Our results suggest that under folding conditions an ensemble of realistically biased protein chains will self-organize into a small number of distinct clusters, each with a large number of thermodynamically, and probably structurally, similar conformers. Together, these several clusters are expected to cover the major, thermodynamically accessible population.

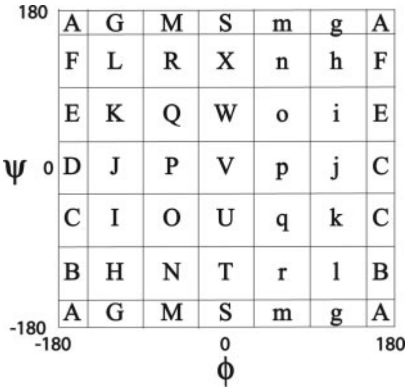
In our approach, ϕ,ψ-space is partitioned into a uniform, labeled grid of 36 squares, each 60° × 60° (Fig. 1), a discretized version of dipeptide map (32). Every square, termed a mesostate, corresponds to a coarse-grained value of a ϕ,ψ-pair, and therefore a linear string of mesostates is a highly approximate description of a protein's 3D structure. By adopting this mesostate representation, our inquiry can be focused into the specific question: can native protein topology be rebuilt solely from the mesostate sequence, without prior knowledge of bond lengths and angles or side-chain identity? This question is of considerable practical interest because approximate

Conflict of interest statement: No conflicts declared.

Abbreviation: rmsd, rms deviation.

*To whom correspondence should be addressed. E-mail: grose@jhu.edu.

www.pnas.org/cgi/doi/10.1073/pnas.0508415102



A	G	M	S	m	g	A
F	L	R	X	n	h	F
E	K	Q	W	o	i	E
D	J	P	V	p	j	C
C	I	O	U	q	k	C
B	H	N	T	r	l	B
A	G	M	S	m	g	A

Fig. 1. Backbone ϕ,ψ-space for a dipeptide was subdivided into 36 alphabetically labeled, 60° × 60° grid squares, called mesostates. A residue's mesostate is a very coarse-grained representation of its backbone conformation (ref. 26 and Table 1).

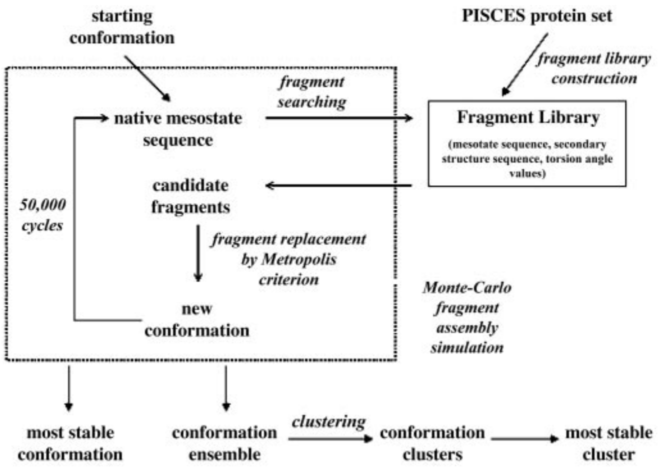


Fig. 2. Flowchart of individual steps (fragment searching and replacement, structure generation, evaluation, and clustering) as described in *Methods*.

torsion angles can be obtained directly from NMR spectroscopy in the form of chemical shifts and residual dipolar couplings (33, 34).

Methods

Even with accurate backbone torsion angles (ϕ and ψ), rebuilding a protein in three dimensions is a nontrivial task (35), owing to the small, but cumulative, deviations from ideality in backbone planarity (i.e., ω-angles), bond lengths, and scalar angles. The method used here is outlined in Fig. 2 and individual steps are described below.

Fragment Library Construction. A protein list of 4,341 chains with sequence identity <40%, resolution >2.5 Å, and R factor of 20 or better was downloaded from the PISCES server (36) and split into all possible overlapping five-residue fragments, 873,352 in all. Fragment lists sorted by backbone torsion angles, mesostate sequence, and secondary structure sequence were generated and stored for later use. Secondary structure assignments were made by using PROSS (26), which maps backbone dihedral angles into mesostates (Fig. 1) and then assigns each residue within a sequence of mesostates to one of five secondary structure categories: T(urn), H(elix), E(xtended), P(olyprolineII), or C(oil).

Generation of 3D coordinates from a mesostate sequence is an iterative process of selecting suitable candidate fragments and assembling them into a coherent structure. For each protein under consideration, a total of 25,000 viable structures was generated in each of 10 parallel simulations, followed by clustering and analysis of a randomly selected representative subset.

Fragment Replacement Criteria. ***Starting conformation.*** Initially, a fully extended polypeptide chain with ideal bond lengths and angles was constructed by using the ribosome package in LINUS (37). Side chains were omitted; residues were modeled as alanine unless the given mesostate was sterically inaccessible to alanine, in which case it was modeled as glycine.

Fragment selection. To replace a five-residue fragment within a target mesostate sequence, candidates for substitution were selected from the fragment library. Library fragments with the identical mesostate sequence were given precedence, but when the set of identities was too small (≤10 members), fragments with a similar mesostate sequence were used instead (i.e., at least three of the five corresponding mesostates were identical). Upon selection, eligible candidates were filtered to remove any fragments from homologous protein chains or any fragments having a secondary structure sequence that failed to match the target chain.

Fragment replacement. Five-residue library fragments were selected at random as described. Replacing a target sequence fragment by an eligible library fragment entailed changing the backbone torsion angles of the target sequence, residue by residue. If the mesostates of the library and target residues were identical, the torsion angles of the target residue were replaced by the torsion angles of the corresponding library residue. However, if the library and target residue mesostates differed, mesostate-constrained random values for the target residue were generated so as to preserve the native mesostate. Specifically, replacement torsions were generated by varying the original ϕ,ψ-angles by ±5° for helix and turn residues or ±10° for other residues, and by varying the original ω-angle by ±2° for all residues. Allowing torsion angles to vary in this way can compensate for restraints imposed by the use of ideal bond lengths and angles (35).

Fragment Assembly by Monte Carlo Simulation with Simulated Annealing. A Metropolis Monte Carlo simulation (38) of 50,000 cycles was then performed on this target chain, preceded by 5,000 relaxation cycles, where each cycle consisted of n-4 steps for a chain of length n. At each step, a randomly chosen five-residue segment of the target peptide was replaced by a randomly chosen library fragment, as described above. The relaxation phase allowed the initially extended chain to settle into a mesostate-compatible conformation. The subsequent simulation was divided into 25,000 equilibration cycles followed by 25,000 cycles during which the chain can explore low energy conformations, using the Metropolis-based energy function described next.

Energy Function. The Metropolis criterion was applied by using an energy function with three simple terms: (*i*) steric exclusion (E_{soft_debump}), (*ii*) hydrogen bonding (E_{HB}), and (*iii*) global compaction ($E_{confine}$).

Soft-debump potential E_{soft_debump}. A soft-debump potential was used to capture steric repulsion between two atoms, a and b. E_{soft_debump} is a soft-sphere potential, E_{soft_sphere}, when the interatomic distance between atoms does not exceed the sum of their van der Waals radii (23) and a hard-sphere potential beyond this distance (13). Specifically,

$$E_{soft\_debump}(\mathrm{a,b}) = \begin{cases} E_{soft\_sphere}(\mathrm{a,b}), & d_{\mathrm{a,b}} \le r_{\mathrm{a}} + r_{\mathrm{b}}, \\ 0, & d_{\mathrm{a,b}} > r_{\mathrm{a}} + r_{\mathrm{b}} \end{cases}$$

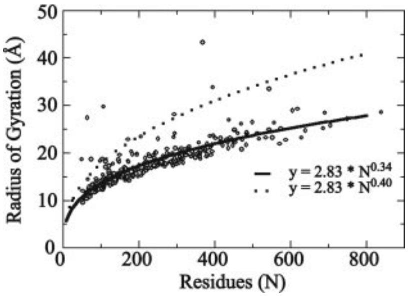


Fig. 3. Two functions of the radius of gyration (*Rg*) vs. protein length (*N*), used as confinement potentials in *Methods*. Functions were calculated as best-fit curves to the observed *Rg* for 337 nonhomologous, x-ray elucidated proteins (○), using the Flory relationship (52), $Rg = R_oN^{\nu}$, as the functional form of the curve. The best-fit curve (solid line) has $\nu = 0.34$, as expected for a self-avoiding polymer in poor solvent. A relaxed function (dashed line) with $\nu = 0.40$ was used in the initial relaxation stage of the simulation.

$$E_{soft\_sphere}(\mathrm{a, b}) = 0.25 \times [(r_{\mathrm{a}} + r_{\mathrm{b}})/d_{\mathrm{a,b}}]^{12},$$

where $d_{\mathrm{a,b}}$ is the distance between atoms a and b and r_{a} and r_{b} are the van der Waals radii for atoms a and b, respectively.

Hydrogen-bond potential E_{HB}. We use an orientation-dependent hydrogen-bond potential (39, 40), with an energy proportional to ε_{HB} when geometric criteria are satisfied (39), and 0 otherwise. A proportionality constant that distinguishes between local (≤4 residues between donor and acceptor) and long-range (>4 residues between donor and acceptor) hydrogen bonds is applied to favor long-range interactions: $\varepsilon_{\mathrm{short}} = 1.0\ \varepsilon_{\mathrm{HB}}$ and $\varepsilon_{\mathrm{long}} = 2.0\ \varepsilon_{\mathrm{HB}}$.

Confinement potential $E_{confine}$. A confinement potential, implemented as a function of radius of gyration, was applied to capture solvent-squeezing effects:

$$E_{confine} = \begin{cases} (Rg - R_0)^2, & Rg \geq R_0, \\ 0, & Rg < R_0 \end{cases}$$

where Rg is the radius of gyration of the current conformation, and R_0 is a threshold value. A value of $R_0 = 2.83 \times N^{0.34}$, close to that predicted by theory (28), was obtained empirically by best-fitting the calculated radius of gyration vs. protein length by using high-resolution crystal structures in the Protein Data Bank (41), as illustrated in Fig. 3. This value was deliberately relaxed to $R_0 = 2.83 \times N^{0.40}$ during the initial 5,000 cycles so as to promote local secondary structure formation.

Table 1. Protein test set

Protein Data Bank ID	Molecule name	Classification	Length, nt
Proteins used in this study			
2GB1	Protein G domain B	α/β	56
1UBQ	Ubiquitin	α/β	76
1C9OA	Cold shock protein, chain A	All β	66
1IFB	Intestinal fatty acid binding protein	All β	131
1VII	Villin head piece	All α	36
1R69	434 Repressor	All α	63
Additional proteins			
1GHHA	DNA damage inducible protein I	α/β	81
1BTB	Barstar	α/β	76
1SHG	α-Spectrin	All β	62
4GCR(1–85)	γ-B crystallin	All β	85
1ENH	Engrailed homeodomain	All α	54
4ICB	Calbindin	All α	76

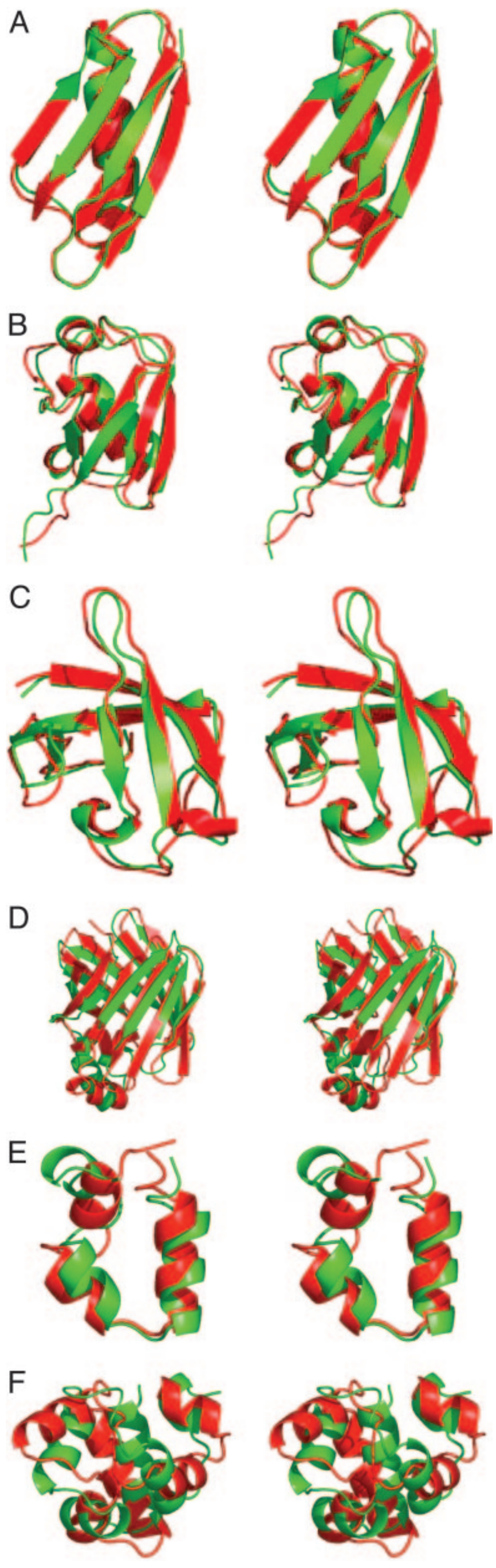


Fig. 4. Stereoviews showing the most stable conformation (red) from simulations superimposed on its corresponding native conformation (green): Protein Data Bank ID codes 2GB1 (*A*), 1UBQ (*B*), 1C9OA (*C*), 1IFB (*D*), 1VII (*E*), and 1R69 (*F*).

Table 2. Backbone rmsd of the most stable conformation

Protein Data Bank ID code	rmsd, Å
2GB1	1.11
1UBQ	1.81
1C90A	1.38
1IFB	3.05
1VII	3.78
1R69	4.49

Clustering. The conformational ensemble for each protein was represented by 500 conformations chosen as every 500th structure from each of the 10 parallel simulations. At this sampling interval, successive structures are expected to be uncorrelated. This ensemble was clustered by using a slightly modified version of CLUSTER 3.0 (42), based on a score that combines the α-carbon distance matrix and differences in torsion angles. Clusters with a similarity coefficient $>70\%$ that spanned $>10\%$ of all conformations were retained for further analysis, and their energy distributions and rms deviation (rmsd) from the native structure were computed.

Our results are insensitive to the particular choice of clustering algorithm. However, the strategy of clustering by structure, not energy, was crucial.

Test Protein Set. The method was tested on multiple proteins. Six representative examples are described here, including two all-α proteins, two all-β proteins, and two α/β-proteins (Table 1). Additional examples not studied here are shown in Table 1. For each protein, the mesostate sequence was obtained by mapping torsion angles into mesostates (Fig. 1), but no other information from the 3D structure was used in these simulations.

Results

For each of the six representative proteins described here, the lowest energy conformation closely resembles its experimentally determined counterpart (Fig. 4). This visual impression is corroborated by the low rmsds between the two corresponding structures (Table 2). Successful simulation was achieved despite the use of coarse-grained backbone angles, the crude energy function, and omission of side chains. Thus, with only these simple energetic criteria, the method was sufficient to discover the native topology in each case listed in Table 1.

Agreement between the lowest energy conformer and the native topology is not merely fortuitous, as is evident from the clustered ensembles in Table 3. For each protein, only a few topologically coherent clusters are sufficient to encompass almost the entire population. In each case, the cluster of lowest energy (bold rows in Table 3) has the lowest rmsd from the native structure. Often, but not invariably, it is also the largest cluster. Regardless, the combination of clustering by structure and then identifying the cluster of lowest energy is sufficient to discover the native topology.

The distributions of simulation energy vs. rmsd were plotted for all 500 sampled conformers in each protein (Fig. 5). Consistent with the observed agreement between energy and native structure in Table 3, these energies either increase monotonically with rmsd or form a single clump, with the sole exception of fatty acid-binding protein (see below). Significantly, the hydrogen-bond potential alone, the only protein-specific term among the three, tracks with the total simulation energy, as seen in Fig. 5.

The energy vs. rmsd distribution for fatty acid-binding protein (Fig. 5) requires special mention. In this case, conformations centered around both 3.0 and 7.0 Å rmsd have similar energies. However, detailed examination of the clusters confirms that both clumps maintain native topology.

Simulations of each protein took between 1 and 2 weeks on a single-processor 2.53-GHz Intel (Santa Clara, CA) Pentium

Table 3. Topological clusters from each ensemble

Protein Data Bank ID code	Cluster	Size*	rmsd, Å†	Energy‡	Hydrogen-bond potentia§
2GBI	**I**	**344**	**1.73 ± 0.36**	**−56.28 ± 2.75**	**−28.40 ± 1.31**
	II	154	2.21 ± 0.37	−56.12 ± 2.84	−28.18 ± 1.35
1UBQ	**I**	**295**	**3.13 ± 0.70**	**−62.00 ± 2.82**	**−31.22 ± 1.39**
	II	74	8.70 ± 0.27	−59.42 ± 3.42	−29.94 ± 1.62
	III	104	5.08 ± 0.44	−60.01 ± 2.01	−30.20 ± 0.89
1C9OA	**I**	**282**	**2.32 ± 1.04**	**−54.63 ± 0.54**	**−27.51 ± 1.76**
	II	50	6.20 ± 0.12	−52.18 ± 1.69	−26.51 ± 0.74
	III	50	10.58 ± 0.34	−43.12 ± 2.59	−21.95 ± 1.69
	IV	50	8.84 ± 0.24	−42.46 ± 1.26	−21.63 ± 0.42
1IFB	I	62	7.12 ± 0.78	−141.08 ± 4.14	−73.56 ± 1.94
	II	188	4.30 ± 0.58	−138.41 ± 3.17	−69.87 ± 1.38
	III	**50**	**2.99 ± 0.05**	**−144.50 ± 1.92**	**−72.38 ± 0.94**
	IV	50	9.17 ± 0.15	−140.95 ± 2.11	−71.49 ± 0.99
1VII	**I**	**400**	**3.72 ± 0.30**	**−34.24 ± 2.59**	**−17.22 ± 1.28**
	II	75	4.10 ± 0.81	−25.84 ± 1.77	−13.06 ± 0.85
1R69	I	98	5.22 ± 0.58	−62.14 ± 2.92	−31.26 ± 1.46
	II	**54**	**4.24 ± 1.12**	**−62.75 ± 3.54**	**−31.54 ± 1.74**
	III	157	5.47 ± 1.05	−60.68 ± 3.26	−30.51 ± 1.60

For each protein, 500 conformers were selected and clustered (500 from each of 10 parallel simulations, as described in *Methods*). Almost all of them can be classified into only a few structurally similar clusters. The lowest-energy cluster is in bold in each case.

*Number of structures in the cluster.

†Average rmsd (±SD) from the native structure for the cluster.

‡Average total simulation energy (±SD) for the cluster.

§Average hydrogen-bond energy (±SD) for the cluster.

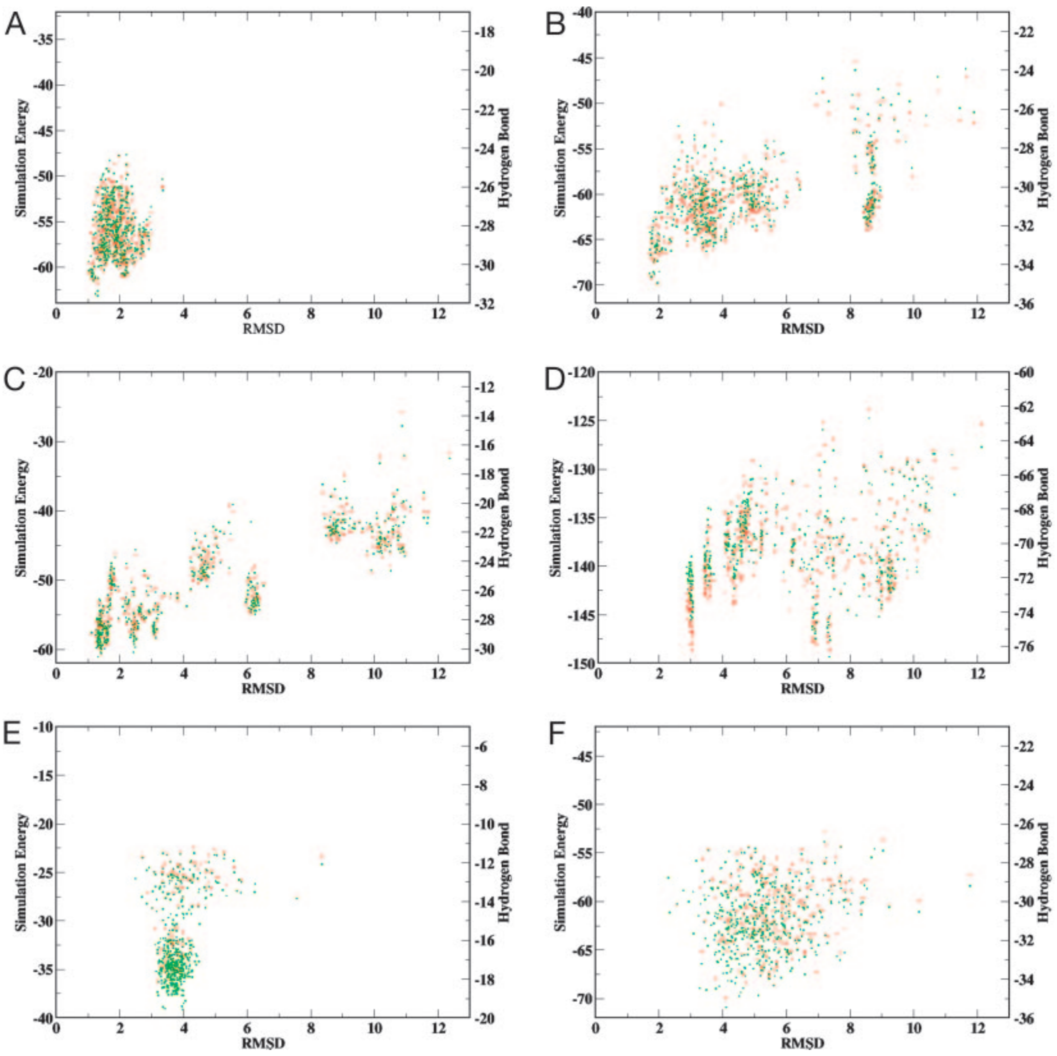


Fig. 5. Simulation energy (red) vs. rmsd from the native structure for 500 conformers in the six simulated ensembles: Protein Data Bank ID codes 2GB1 (*A*), 1UBQ (*B*), 1C9OA (*C*), 1IFB (*D*), 1VII (*E*), and 1R69 (*F*). Notably, the native clump has the lowest energy. Importantly, the energy is dominated by the hydrogen-bond score (green) that tracks with the total simulation energy almost exactly.

processor (i.e., a desktop Unix box) but less than half a day on a 48-node computer farm.

Discussion

Anfinsen's hypothesis (1) established the fundamental relationship between thermodynamics and structure for protein molecules, and since then, the field has sought an energy function that is sufficient to calculate the native conformation. Remarkably, the crude, three-term energy function used here achieves this long-standing goal in simulations of mesostate-constrained conformers. Why?

Our simulations demonstrate that the number of thermodynamically viable clusters consistent with native mesostates is limited. Seemingly, this result is surprising because each ϕ,ψ-pair in a mesostate sequence can assume a large range of values ($60° \times 60°$), suggestive of numerous possible topologies. However, most putative topologies are not feasible, either owing to steric clash (23) or the failure of backbone polar groups to satisfy their hydrogen-bonding requirements (20). Systematic steric clash is quite local, between atoms within six residues of each other in the linear sequence (2), effectively eliminating most topologies that would otherwise be compatible with a specific mesostate sequence (19, 43).

This realization is the basis for the choice of terms in our energy potential: molecular compaction, soft-sphere repulsion, and hydrogen bonding. The first two terms are nonspecific and were chosen to capture the high packing density of proteins (44) while avoiding steric violations (19, 32). Hydrogen bonding is the only protein-specific term, and it plays a central role in protein folding (20, 21, 45) and the dominant role in our simulations (Fig. 5).

With the exception of E_{soft_debump}, the potentials used here are fitted terms that lack meaningful reference energies, and consequently units have been omitted. However, E_{HB}, which tracks with the total energy, is just a count of local and long-range hydrogen bonds, as described in *Methods*, and this quantity is plotted in Fig. 5.

Returning now to the question of how these three criteria are sufficient to identify the native cluster, numerous misfolded conformations having native-like energies are eliminated effectively by restricting the ensemble to conformers with native mesostates. Among those surviving, the lowest energy compact, clash-free, hydrogen-bonded cluster is the native one.

Given the fact that most systematic steric clash is local (2, 23), fragment-assembly Monte Carlo simulation is an attractive strategy for discovering the native topology. Database fragments excised from experimental structures are intrinsically clash-free and locally energy-minimized (43, 46). Accordingly, fragment assembly is the most effective current method for protein structure prediction (47), and it has been used with impressive success in the ROSETTA program, developed by Baker and colleagues (48–51).

In summary, for the six proteins studied here, our fragment-assembly Monte Carlo algorithm successfully identified the native backbone topology solely from its mesostate sequence. Presumably, the addition of side chains would further bias the distribution toward the native topology. In all cases, that topology corresponds to the lowest energy structure, using a simple, three-term potential involving molecular compaction, steric repulsion, and hydrogen bonding. Remarkably, the hydrogen-bond potential alone closely tracks the total energy. Our algorithm was developed with an eye toward the practical problem of solving 3D structures from NMR data by mapping chemical shifts and/or residual dipolar couplings onto mesostates (33, 34). Initial results indicate that ubiquitin can be solved in this way (data not shown).

We thank Nicholas Fitzkee, Nicholas Panasik, Robert Schleif, and Timothy Street for useful suggestions. Support from the Burroughs Wellcome Fund (to H.G.) and the Mathers Foundation (to G.D.R.) is gratefully acknowledged.

1. Anfinsen, C. B. (1973) *Science* **181,** 223–230.
2. Fitzkee, N. C. & Rose, G. D. (2005) *J. Mol. Biol.*, in press.
3. DePristo, M. A., de Bakker, P. I. & Blundell, T. L. (2004) *Structure (London)* **12,** 831–838.
4. Fitzkee, N. C., Fleming, P. J., Gong, H., Panasik, N., Jr., Street, T. O. & Rose, G. D. (2005) *Trends Biochem. Sci.* **30,** 73–80.
5. Shortle, D. (2002) *Adv. Protein Chem.* **62,** 1–23.
6. Shortle, D. & Ackerman, M. S. (2001) *Science* **293,** 487–489.
7. Louhivuori, M., Paakkonen, K., Fredriksson, K., Permi, P., Lounila, J. & Annila, A. (2003) *J. Am. Chem. Soc.* **125,** 15647–15650.
8. Jha, A. K., Colubri, A., Freed, K. F. & Sosnick, T. R. (2005) *Proc. Natl. Acad. Sci. USA* **102,** 13099–13104.
9. Sallum, C. O., Martel, D. M., Fournier, R. S., Matousek, W. M. & Alexandrescu, A. T. (2005) *Biochemistry* **44,** 6392–6403.
10. Tiffany, M. L. & Krimm, S. (1968) *Biopolymers* **6,** 1767–1770.
11. Rucker, A. L. & Creamer, T. P. (2002) *Protein Sci.* **11,** 980–985.
12. Creamer, T. P. & Campbell, M. N. (2002) *Adv. Protein Chem.* **62,** 263–282.
13. Pappu, R. V. & Rose, G. D. (2002) *Protein Sci.* **11,** 2437–2455.
14. Mezei, M., Fleming, P. J., Srinivasan, R. & Rose, G. D. (2004) *Proteins* **55,** 502–507.
15. Shi, Z., Olson, C. A., Rose, G. D., Baldwin, R. L. & Kallenbach, N. R. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 9190–9195.
16. Shi, Z., Woody, R. W. & Kallenbach, N. R. (2002) *Adv. Protein Chem.* **62,** 163–240.
17. Tran, H. T., Wang, X. & Pappu, R. V. (2005) *Biochemistry*, **44,** 11369–11380.
18. Jha, A. K., Colubri, A., Zaman, M. H., Koide, S., Sosnick, T. R. & Freed, K. F. (2005) *Biochemistry* **44,** 9691–9702.
19. Fitzkee, N. C. & Rose, G. D. (2004) *Protein Sci.* **13,** 633–639.
20. Fleming, P. J. & Rose, G. D. (2005) *Protein Sci.* **14,** 1911–1917.
21. Panasik, N., Jr., Fleming, P. J. & Rose, G. D. (2005) *Protein Sci.*, in press.
22. Baldwin, R. L. & Zimm, B. H. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 12391–12392.
23. Pappu, R. V., Srinivasan, R. & Rose, G. D. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 12565–12570.
24. Fitzkee, N. C., Fleming, P. J. & Rose, G. D. (2005) *Proteins* **58,** 852–854.
25. Swindells, M. B., MacArthur, M. W. & Thornton, J. M. (1995) *Nat. Struct. Biol.* **2,** 596–603.
26. Srinivasan, R. & Rose, G. D. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 14258–14263.
27. Kohn, J. E., Millett, I. S., Jacob, J., Zagrovic, B., Dillon, T. M., Cingel, N., Dothager, R. S., Seifert, S., Thiyagarajan, P., Sosnick, T. R., *et al.* (2004) *Proc. Natl. Acad. Sci. USA* **101,** 12491–12496.
28. Fleming, P. J. & Rose, G. D. (2005) in *Protein Folding Handbook*, ed. Buchner, T. K. A. J. (Wiley-VCH, Weinheim, Germany), Part 1, Vol. 2, pp. 710–736.
29. Fitzkee, N. C. & Rose, G. D. (2004) *Proc. Natl. Acad. Sci. USA* **101,** 12497–12502.
30. Cohen, F. E., Richmond, T. J. & Richards, F. M. (1979) *J. Mol. Biol.* **132,** 275–288.
31. Kihara, D. (2005) *Protein Sci.* **14,** 1955–1963.
32. Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. (1963) *J. Mol. Biol.* **7,** 95–99.
33. Cornilescu, G., Delaglio, F. & Bax, A. (1999) *J. Biomol. NMR* **13,** 289–302.
34. Chou, J. J., Li, S. & Bax, A. (2000) *J. Biomol. NMR* **18,** 217–227.
35. Holmes, J. B. & Tsai, J. (2004) *Protein Sci.* **13,** 1636–1650.
36. Wang, G. & Dunbrack, R. L., Jr. (2003) *Bioinformatics* **19,** 1589–1591.
37. Srinivasan, R., Fleming, P. J. & Rose, G. D. (2004) *Methods Enzymol.* **383,** 48–66.
38. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953) *J. Chem. Phys.* **21,** 1087–1092.
39. Kortemme, T., Morozov, A. V. & Baker, D. (2003) *J. Mol. Biol.* **326,** 1239–1259.
40. Morozov, A. V., Kortemme, T., Tsemekhman, K. & Baker, D. (2004) *Proc. Natl. Acad. Sci. USA* **101,** 6946–6951.
41. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000) *Nucleic Acids Res.* **28,** 235–242.
42. de Hoon, M. J. L., Imoto, S., Nolan, J. & Miyano, S. (2004) *Bioinformatics* **20,** 1453–1454.
43. Fujitsuka, Y., Takada, S., Luthey-Schulten, Z. A. & Wolynes, P. G. (2004) *Proteins* **54,** 88–103.
44. Richards, F. M. (1977) *Annu. Rev. Biophys. Bioeng.* **6,** 151–176.
45. Baldwin, R. L. (2003) *J. Biol. Chem.* **278,** 17581–17588.
46. Lee, J., Kim, S.-Y., Joo, K., Kim, I. K. & Lee, J. (2004) *Proteins* **56,** 704–714.
47. Chikenji, G., Fujitsuka, Y. & Takada, S. (2003) *J. Chem. Phys.* **119,** 6895–6903.
48. Bonneau, R., Tsai, J., Ruczinski, I., Chivian, D., Rohl, C., Strauss, C. E. M. & Baker, D. (2001) *Proteins* **45,** 119–126.
49. Bonneau, R., Strauss, C. E. M., Rohl, C. A., Chivian, D., Bradley, P., Malmstrom, L., Robertson, T. & Baker, D. (2002) *J. Mol. Biol.* **322,** 65–78.
50. Simons, K. T., Ruczinski, I., Kooperberg, C., Fox, B. A., Bystroff, C. & Baker, D. (1999) *Proteins* **34,** 82–95.
51. Simons, K. T., Kooperberg, C., Huang, E. & Baker, D. (1997) *J. Mol. Biol.* **268,** 209–225.
52. Flory, P. J. (1969) *Statistical Mechanics of Chain Molecules* (Wiley, New York).