PROTEINS: Structure, Function, and Bioinformatics 61:338–343 (2005)

Does Secondary Structure Determine Tertiary Structure in Proteins?

Haipeng Gong and George D. Rose*
Jenkins Department of Biophysics, Johns Hopkins University, Baltimore, Maryland

ABSTRACT **Is highly approximate knowledge of a protein's backbone structure sufficient to successfully identify its family, superfamily, and tertiary fold? To explore this question, backbone dihedral angles were extracted from the known three-dimensional structure of 2,439 proteins and mapped into 36 labeled, 60° × 60° bins, called *mesostates*. Using this coarse-grained mapping, protein conformation can be approximated by a linear sequence of mesostates. These linear strings can then be aligned and assessed by conventional sequence-comparison methods. We report that the mesostate sequence is sufficient to recognize a protein's family, superfamily, and fold with good fidelity. Proteins 2005;61:338–343.**

INTRODUCTION

Elements of local secondary structure – helices, hairpins and turns – form on a fast timescale.[1,2] In the hierarchic model of folding,[3–5] these local elements then further organize to form tertiary structure.[6–8] How many ways can a given set of local structural elements be assembled into a coherent three-dimensional structure? Many ways? Several ways? Uniquely?

In globular proteins, primary structure determines tertiary structure uniquely,[9] and based on this tenet, computer programs such as BLAST[10] and FASTA[11] have been designed to recognize protein homologs in a database, using pairwise sequence alignment. Although similar sequence implies similar structure,[12,13] the converse proposition does not hold: proteins with the same fold in the Structure Classification of Proteins (SCOP) database[14] may have dissimilar sequences. Specifically, as the aligned sequence identity between homologous proteins falls below ~30%, successful detection by pairwise alignment diminishes markedly.[12,15] In response, newer sequence-to-profile and profile-to-profile methods[16,17] have been developed that boost detection sensitivity for distantly related proteins having similar structures. Such methods are still based on sequence comparisons. Recently, however, Honig and coworkers introduced a profile-to-profile alignment program[18] that improves the performance of remote homology detection by combining secondary structure information with primary sequence.

In fact, the work of Przytycka et al.[19] suggests that secondary structure alone may be sufficient to recognize tertiary structure. In their study, 183 proteins, with less than 30% aligned sequence identity,[20] were represented as strings of secondary structure elements, including turns and loops. Using a simple scoring matrix, conventional pairwise sequence comparisons between these strings were performed and used to construct a *Przytycka-tree* (P-tree), in which the distance between any two nodes is proportional to the difference in score between their aligned secondary structure strings. The P-tree is generated completely automatically, and it reflects the global secondary structure relationships among the proteins used to construct it: the closer the nodes, the greater the similarity of secondary structure among their corresponding proteins. Surprisingly, the straightforward P-tree was found to be largely in agreement with the SCOP tree, although the latter is a complex construct based on structure, evolutionary knowledge, and human judgment. This result lends support to the hypothesis that successful fold recognition can be derived solely from knowledge of secondary structure.

We extend this idea here by quantifying the degree to which approximate backbone conformation can determine the protein fold. This question is of considerable practical interest in that such information can often be obtained directly from experiment, using residual dipolar coupling,[21] for example.

In brief, our approach involves subdividing backbone ϕ,ψ-space into 36 labeled, 60° × 60° grid squares, called mesostates (Fig. 1). Backbone dihedral angles were extracted from proteins of known structure and used to map each residue into its corresponding mesostate, after which the protein was rewritten as a linear string of mesostates. Pairwise alignment of mesostate sequences was performed using dynamic programming, with a mesostate substitution matrix derived from simulation.[22] Our approach is related to the earlier work of Sali and Blundell[23] and of Eisenberg and coworkers.[24]

SCOP family, superfamily, and fold benchmark tests[14] were used for comparison with the following:

(1) established pairwise alignment programs WU-BLAST, FASTA, and Smith–Waterman,[25]
(2) a sequence-to-profile alignment program, SAM-T2K,[17] and
(3) a structure alignment program, VAST (Vector Align-

*Correspondence to: George D. Rose, T. C. Jenkins Department of Biophysics, The Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218-2608. Email: grose@jhu.edu

Received 24 February 2005; Accepted 8 April 2005

Published online 15 August 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20622

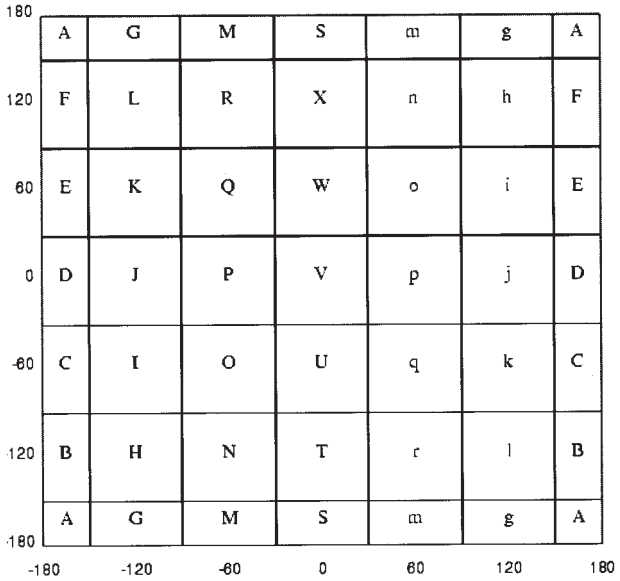

Fig. 1. Dihedral angle space for a dipeptide, partitioned into 36 bins, each corresponding to one mesostate.

ment Search Tool),[26] which is based solely on three-dimensional information.

Results indicate that our approach performed quite well for family recognition, and satisfactorily in fold and superfamily recognition.

MATERIALS AND METHODS

Dynamic Programming

A Needleman–Wunsch algorithm,[27] MESO_ALIGN, was developed to perform global alignment between two mesostate strings, with allowance for a gap penalty. The primary alignment score between a pair of mesostates, $V(i,j)$, is the sum of two separate components:

$$V(i,j) = \mathrm{MESO}(i,j) + \mathrm{SS}(i,j) \tag{1}$$

where $\mathrm{MESO}(i,j)$ represents the substitution score obtained from a mesostate substitution matrix, and $\mathrm{SS}(i,j)$ represents the substitution score obtained from a companion secondary-structure substitution matrix.

The gap penalty, Gap, is also composed of two parts, a conventional affine gap penalty and an additional penalty when residues with repetitive secondary structure are aligned with a gap:

$$\mathrm{Gap} = \mathrm{affine\_gap} + \mathrm{additional\_gap} \tag{2}$$

Differing affine gap penalties are assigned to central and flanking residues in a sequence of interest:

$$\mathrm{affine\_gap} = \begin{cases} 0, \text{ for flanking residues} \\ 20 + 2 \times l, \text{ for central residues} \end{cases} \tag{3}$$

$$\mathrm{additional\_gap} = l \times n_{\mathrm{SS}} \tag{4}$$

where n_{SS} is the number of residues with repetitive secondary structure in the gap, and l is the length of the gap. The final score is the alignment score divided by the length of the aligned sequence, including gaps.

Substitution Matrices

The mesostate substitution matrix was constructed from hard-sphere Monte Carlo simulations of a nine-residue polyala peptide.[28] Five simulations were run in parallel, with each of the 10,000 cycles preceded by 500 equilibration cycles. The mesostate transition frequencies, $f(i,j)$, between any two mesostates i and j were monitored for the central residue, Ala_5, and averaged over the five simulations. To obtain a symmetric matrix, $f(i,j)$ and $f(j,i)$ were assigned their mean when the two values differed. The similarity score was then calculated as:

$$\mathrm{MESO}(i,j) = \left(10 \times \log \frac{f(i,j)}{f(i) \times f(j)}\right)_{\mathrm{int}} \tag{5}$$

where $f(i)$ is the total single mesostate occupation frequency, and 10 is a scaling factor introduced for convenience.

The secondary-structure substitution matrix was constructed subjectively, not calculated. Its elements were deliberately set to have smaller values than those of the mesostate substitution matrix to ensure that the secondary structure score would play a subordinate role in the alignment.

Mesostate and Secondary Structure Assignment

A mesostate assignment was made from three-dimensional coordinates by calculating the backbone dihedral angles for each residue and mapping them as indicated in Figure 1. In turn, the mesostate sequence was converted into one of four secondary structure assignments using PROSS[29] (available from www.roselab.jhu.edu): T(urn), H(elix), E(xtended), or C(oil). In detail, β-turns (T) were assigned to two-residue segments having mesostates from the set {OO, OP, OJ, PO, PP, PJ, JO, JP, JJ, Mo, Mp, Mj, Ro, Rp, Rj, oo, op, oj, po, pp, pj, jo, jp, jj, mO, mP, mJ, rO, rP, rJ}. α-Helices (H) were assigned to segments of five or more residues having mesostates from the set {O, P}. β-Strands (E) were assigned to segments of three or more successive residues having mesostates from the set {L, G, F, A, R, M}. Residues not classified into one of these three categories were assigned as coil (C). Finally, the secondary structure sequence was adjusted using the following three rules:

(1) A single C residue flanked by helical residues (H) was converted to H.
(2) A single C residue flanked by strand residues (E) was converted to E.
(3) Sequences with more than five consecutive turn residues (T) were converted to H if all mesostates were from the set {J, O, P}.

SCOP Benchmark Test

SCOP domains (version 1.63) with less than 40% sequence identity were obtained from the ASTRAL SCOP

TABLE I. Programs Used in SCOP Benchmark Tests

Name	Package version	Source	Parameter
FASTA	FASTA3.3	ftp://ftp.virginia.edu/pub/fasta/	Gap = $14 + 2 \times n$
SMITH–WATERMAN	FASTA3.3	ftp://ftp.virginia.edu/pub/fasta/	Gap = $14 + 2 \times n$
WU-BLAST	WU-BLAST2.0	http://blast.wustl.edu/blast/executables/	Gap = $9 + 2 \times n$
SAM	SAM-T2K	http://www.cse.ucsc.edu/research/compbio/sam.html	Filter: WU-BLAST2.0
VAST	VAST	ftp://ftp.ncbi.nih.gov/mmdb/vastdata/	P-value cutoff: 0, 4

Database[30] (astral.stanford.edu) and filtered to retain only structures with resolution ≤2.0 Å. Domains comprised of more than one chain or fewer than four elements of secondary structure were also filtered. After filtering, 2439 domains remained in our library. The domain having the best resolution was chosen from each family with at least three family members, resulting in 113 domains for use in the SCOP family benchmark test set. Applying the same criteria, 144 and 159 domains were identified for the SCOP superfamily and fold benchmark test sets, respectively.

The programs used for comparison included FASTA, SMITH–WATERMAN, SAM-T2K, WU-BLAST, and VAST (Table I). Default parameters were used in all cases except that two different p-value cutoffs (0 and 4) were applied separately for VAST. The BLOSUM62 matrix[31] was adopted for all sequence-alignment programs. All programs allow self-detection.

Global alignment algorithms do not follow an underlying probability distribution, and therefore empirical estimates, in the form of z-scores, were used to gauge alignment similarity. To obtain a mean and variance for each query domain, a random library with the same length distribution as our benchmarks was constructed by randomly connecting six-residue segments taken from the test sets. For every query domain, the programs MESO_ALIGN, FASTA, SMITH–WATERMAN, and SAM-T2K were scored against all domains in the test set. Z-scores were then calculated from the mean and variance obtained by running the same program on the corresponding random library instead. Specificity (accuracy) and sensitivity (coverage) were computed as a function of threshold using the formulae:

$$\text{Specificity} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100\% \quad (6)$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{TN}} \times 100\% \quad (7)$$

where TP (true positive) is the number of proteins that were correctly identified as belonging to the same class (i.e. family, superfamily, or fold) as the query, FP (false positive) is the number of proteins that were incorrectly identified as belonging to the same class as the query, and TN (true negative) is the number of proteins that were incorrectly identified as belonging to a different class than the query. Defined in this way, *specificity* measures the percentage of correct predictions, while *sensitivity* measures the fraction of the library covered by the predictions. Using these data, "specificity vs. sensitivity" curves were drawn for each program. The mean and variance are not easily obtained for VAST and WU-BLAST. Instead, the score normalized by the length of the query domain (VAST) and the negative logarithm of expectation (WU-BLAST) were used in lieu of the Z-score for these two programs.

	G	J	L	M	O	P	R	o	p	R	Z
G	8	0	4	3	-5	-2	2	1	0	-1	-2
J	0	10	0	1	-4	-4	-1	-2	-1	0	-2
L	4	0	7	3	-4	-3	2	1	-1	0	-2
M	3	1	3	10	-4	-3	2	0	0	1	-2
O	-5	-4	-4	-4	2	-4	-4	-5	-5	-4	-2
P	-2	-4	-3	-3	-4	6	-3	-3	-4	-3	-2
R	2	-1	2	2	-4	-3	10	0	-1	0	-2
O	1	-2	1	0	-5	-3	0	9	-1	-1	-2
P	0	-1	-1	0	-5	-4	-1	-1	10	-1	-2
R	-1	0	0	1	-4	-3	0	-1	-1	10	-2
Z	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2	-1

	H	E	T	C
H	2	-2	-1	0
E	-2	2	-2	0
T	-1	-2	2	0
C	0	0	0	1

Fig. 2. The (**A**) mesostate and (**B**) secondary-structure substitution matrices.

RESULTS AND DISCUSSION

Substitution Matrices

The mesostate and the secondary-structure substitution matrices are shown in Figure 2. Elements of the mesostate substitution matrix are similarity scores between any two accessible mesostates as described in Materials and Methods. In physical terms, this score can be interpreted as the relative height of the energy barrier between any two states; transitions between dissimilar states (e.g., helical and extended) will then have negative scores. Only the 10 accessible mesostates are listed in the substitution matrix; all others are represented by "Z." Accessible mesostates fall into three groups: right-handed helical {O, P, J}, extended {L, G, M, R}, and left-handed helical {o, p, r}. In combination, these 10 mesostates can adopt almost any

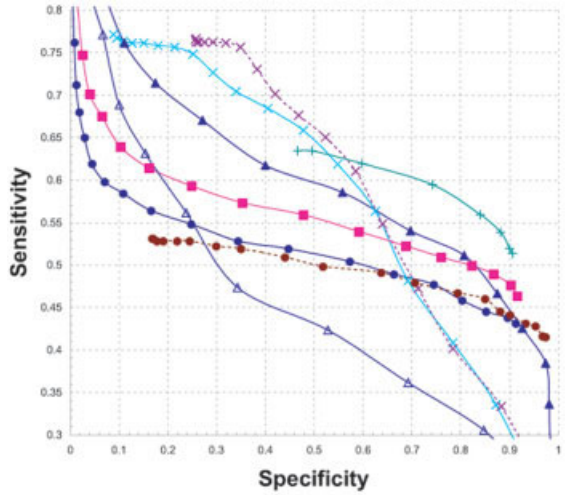

Fig. 3. Specificity vs. sensitivity curve of SCOP family benchmark test. Programs include: MESO_ALIGN (solid light blue line with triangles); SMITH–WATERMAN (solid pink line with squares); FASTA (solid dark blue line with circles); WU-BLAST (dashed red line with circles); SAM (solid aqua line with pluses); VAST with p-value cutoff = 0 (solid cyan line with forks); VAST with p-value cutoff = 4 (dashed purple line with forks). As a control, secondary structure alignment is also shown (solid blue line with open triangles).

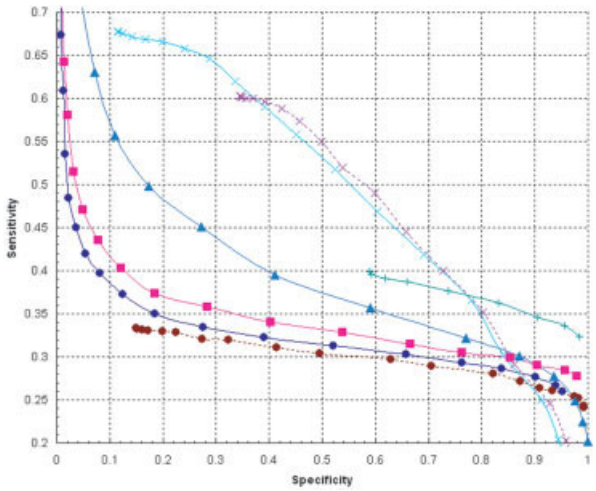

Fig. 4. Specificity vs. sensitivity curve of SCOP superfamily benchmark test. Programs and their annotations are as in Figure 3.

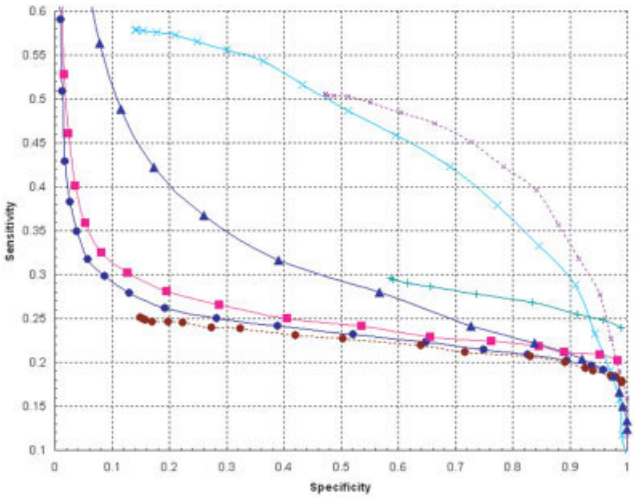

Fig. 5. Specificity vs. sensitivity curve of SCOP fold benchmark test. Programs and their annotations are as in Figure 3.

secondary structure. Exceptions involve structures with mesostates from {A, F, m, j} that are found only rarely in the Protein Data Bank (PDB) or are accessible only to glycine.

The secondary-structure substitution matrix was constructed to reward correct alignment between elements of secondary structure, and, conversely, to penalize their misalignment. Similar to the strategy used by Przytycka et al.,[19] the penalty is highest for attempted alignment between helix/turn and extended structure. Matrix values were chosen to be smaller than scores in the mesostate substitution matrix, such that the latter play the dominant role in aligning structures [in eq. (1)].

Benchmark Tests

Three benchmark tests are reported as "specificity vs. sensitivity" curves for the SCOP family, superfamily, and fold benchmarks, respectively (Figs. 3–5). Our mesostate alignment algorithm performs quite well in family recognition (Fig. 3). Additionally, there is improved performance over some previous methods[18] in both fold and superfamily tests. Possibly, these ostensible improvements are merely for technical reasons: i.e. self-detection is allowed in our benchmark tests, and representatives in our test set libraries are limited to those with at least two other library members in each case. Nevertheless, the relative position of the MESO_ALIGN curves exceeds that of the pairwise sequence alignment programs (FASTA, WU-BLAST, and SMITH–WATERMAN) in all three benchmarks. SAM, a sequence-to-profile method, also exceeds pairwise alignment in all benchmarks.

Summarizing, MESO_ALIGN offers improved performance over pairwise sequence alignment programs. Indeed, cases can be identified in which MESO_ALIGN outperforms all other methods tested here; three such examples are given in Table II. In the family benchmark test (Fig. 3), its performance rivals VAST in the middle range and even exceeds VAST (as well as the other pairwise alignment methods) at high specificity (≥0.6), suggesting that mesostate information may be sufficient to determine three-dimensional structure.

Discussion

Despite the use of coarse-grained mesostates, crude matrices, a simple gap penalty, and a naïve pairwise alignment algorithm, the effective performance of MESO_ALIGN in family recognition invites the conjecture that secondary-structure information alone is sufficient to determine the overall tertiary fold. If true, the NP-complete problem of aligning two protein structures can be simplified to the polynomial-time problem of aligning two one-dimensional secondary structure sequences.

TABLE II. Examples of Structural Similarities Identified by Meso_Align But Not by Other Methods

SCOP ID	PDB ID Chain residues	RANK					
		MESO_ALIGN	FASTA	BLAST	Smith–Waterman	VAST	SAM
a.25.1.2	1KGN A: 2–297	2	1173	***	1371	+++	45
b.60.1.1	1LF7 A: 10–180	4	2360	***	2180	***	661
c.55.1.1	1BUP A: 4–188	3	1274	***	1231	+++	346

*** not found.
+++ found with P-value cutoff = 0 but not found with P-value cutoff = 4.

It is often assumed that structurally related proteins in the PDB are Boltzmann-distributed over a narrow energy range[32] that has attained equilibrium during the time-course of evolution. Accordingly, the score in a PAM matrix can be interpreted as representing the free energy cost of residue substitution[33] in such a process. Similarly, the scores in our mesostate substitution matrix can also be regarded as the free energy cost of transitions between mesostate pairs during protein folding, given that they were obtained from Monte Carlo simulation[29] using the Metropolis criterion.[34] Furthermore, the total alignment score would then reflect the energy cost, relative to the self-alignment, of transforming one backbone conformation into another. Thus, an advantage of mesostate alignment is that it can detect thermodynamically interconvertible structures, at least in principle.

Another distinct advantage of shifting the focus to thermodynamically interconvertible structures is their lack of dependence on a global coordinate system. For example, altering a flexible loop between two protein domains could confound structural comparison methods such as VAST. However, the original and altered conformations would still be detected as thermodynamically interconvertible by mesostate alignment.

The approach described in this article can be regarded as an ongoing experiment, with opportunities for improvement. For example, our substitution matrix was derived from a rudimentary simulation and can undoubtedly be further refined. Other strategies for devising the matrix could also be explored. Strategies based on evolutionary arguments are especially appealing, such as generating the matrix from aligned homologs, like a PAM matrix[33] or from aligned key segments, like a BLOSUM matrix.[31]

Sequence information was deliberately neglected in this work because we wanted to test the degree to which secondary structure alone is sufficient for fold and family recognition. However, side chains play an important role in the folding process,[35] and their inclusion would complement and presumably improve detection sensitivity and specificity; see for example ref. 23. The inclusion of sequence information might result in apparent improvement for a technical reason as well: some SCOP assignments are based not only on structure similarity but also on sequence conservation at key positions. In such cases, the SCOP "gold standard" is not entirely compatible with algorithms based solely on backbone structure (e.g., VAST and MESO_ALIGN).

A final strategy for improvement involves reformulating the mesostate alignment program as a sequence-to-profile or a profile-to-profile algorithm. It was shown that comparisons using multiple sequences detect three times as many remote homologs as pairwise methods.[15] A similar finding is evident in Figures 3–5: the performance of SAM, a sequence-to-profile algorithm, exceeds that of SMITH–WATERMAN, a pairwise method. Recent profile-to-profile algorithms enhance performance even further.

ACKNOWLEDGMENTS

We thank Dr. Steve Bryant for help with VAST and Nicholas Fitzkee, Patrick Fleming, Nicholas Panasik, and Timothy Street for numerous suggestions. Support from the Burroughs Wellcome Fund (HG) and the Mathers Foundation (GDR) is gratefully acknowledged.

REFERENCES

1. Muñoz V, Thompson PA, Hofrichter J, Eaton WA. Folding dynamics and mechanisms of beta-hairpin formation. Nature 1997;390:196–199.
2. Gong H, Isom DG, Srinivasan R, Rose GD. Local secondary structure content predicts folding rates for simple, two-state proteins. J Mol Biol 2003;327(5):1149–1154.
3. Rose GD. Hierarchic organization of domains in globular proteins. J Mol Biol 1979;134:447–470.
4. Baldwin RL, Rose GD. Is protein folding hierarchic? II. Folding intermediates and transition states. Trends Biochem Sci 1999;24(2):77–83.
5. Baldwin RL, Rose GD. Is protein folding hierarchic? I. Local structure and peptide folding. Trends Biochem Sci 1999;24(1):26–33.
6. Karplus M, Weaver DL. Diffusion-collision model for protein folding. Biopolymers 1979;18:1421–1437.
7. Pappu RV, Weaver DL. The early folding kinetics of apomyoglobin. Protein Sci 1998;7(2):480–490.
8. Myers JK, Oas TG. Preorganized secondary structure as an important determinant of fast protein folding. Nat Struct Biol 2001;8(6):552–558.
9. Anfinsen CB. Principles that govern the folding of protein chains. Science 1973;181(96):223–230.
10. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990;215(3):403–410.
11. Pearson WR, Lipman DJ. Improved tools for biological sequence comparison. Proc Natl Acad Sci USA 1988;85(8):2444–2448.
12. Doolittle RF. Of urfs and orfs. Mill Valley, CA: University Science Books; 1986.
13. Sander C, Schneider R. Database of homology-derived protein structures and the structural meaning of sequence alignment. Proteins 1991;9(1):56–68.
14. Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 1995;247(4):536–540.
15. Park J, Karplus K, Barrett C, Hughey R, Haussler D, Hubbard T, Chothia C. Sequence comparisons using multiple sequences detect three times as many remote homologues as pairwise methods. J Mol Biol 1998;284(4):1201–1210.
16. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25(17):3389–3402.

17. Karplus K, Sjolander K, Barrett C, Cline M, Haussler D, Hughey R, Holm L, Sander C. Predicting protein structure using hidden Markov models. Proteins Suppl 1997;1:134–139.
18. Tang CL, Xie L, Koh IY, Posy S, Alexov E, Honig B. On the role of structural information in remote homology detection and sequence alignment: new methods using hybrid sequence profiles. J Mol Biol 2003;334(5):1043–1062.
19. Przytycka T, Aurora R, Rose GD. A protein taxonomy based on secondary structure. Nat Struct Biol 1999;6(7):672–682.
20. Hobohm U, Sander C. Enlarged representative set of protein structures. Protein Science 1994;3(3):522–524.
21. Zweckstetter M, Bax A. Single-step determination of protein substructures using dipolar couplings: aid to structural genomics. J Am Chem Soc 2001;123(38):9490-9491.
22. Srinivasan R, Rose GD. Ab initio prediction of protein structure using LINUS. Proteins 2002;47(4):489–495.
23. Sali A, Blundell TL. Definition of general topological equivalence in protein structures. A procedure involving comparison of properties and relationships through simulated annealing and dynamic programming. J Mol Biol 1990;212(2):403–428.
24. Luthy R, Bowie JU, Eisenberg D. Assessment of protein models with three-dimensional profiles. Nature 1992;356(6364):83–85.
25. Smith TF, Waterman MS. Identification of common molecular subsequences. J Mol Biol 1981;147(1):195–197.
26. Gibrat JF, Madej T, Bryant SH. Surprising similarities in structure comparison. Curr Opin Struct Biol 1996;6(3):377–385.
27. Needleman SB, Wunsch CD. A general method applicable to the search for similarities in the amino acid sequence of two proteins. J Mol Biol 1970;48(3):443–453.
28. Srinivasan R, Rose GD. Ab initio protein folding using LINUS. Methods Enzymol 2004;383:48–66.
29. Srinivasan R, Rose GD. A physical basis for protein secondary structure. Proc Natl Acad Sci USA 1999;96(25):14258–14263.
30. Brenner SE, Koehl P, Levitt M. The ASTRAL compendium for protein structure and sequence analysis. Nucleic Acids Res 2000; 28(1);254–256.
31. Henikoff S, Henikoff JG. Amino acid substitution matrices from protein blocks. Proc Natl Acad Sci USA 1992;89:10915–10919.
32. Shortle D. Propensities, probabilities, and the Boltzmann hypothesis. Protein Sci 2003;12(6):1298–1302.
33. Dayhoff MO, Schwartz RM, Orcutt BC. A model of evolutionary change in proteins. In: Dayhoff MO, editor. Atlas of protein sequence and structure, vol. 5. Silver Spring, MD: National Biomedical Research Foundation; 1978. p 345–352.
34. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. J Chem Phys 1953;21:1087–1092.
35. Frieden C. The kinetics of side chain stabilization during protein folding. Biochemistry 2003;42(43):12439–12446.