Protein Structure Prediction by Threading Methods: Evaluation of Current Techniques

Christian M.-R. Lemer, Marianne J. Rooman, and Shoshana J. Wodak
Unité de Conformation de Macromolécules Biologiques, 1050 Brussels, Belgium

ABSTRACT **This paper evaluates the results of a protein structure prediction contest. The predictions were made using threading procedures, which employ techniques for aligning sequences with 3D structures to select the correct fold of a given sequence from a set of alternatives. Nine different teams submitted 86 predictions, on a total of 21 target proteins with little or no sequence homology to proteins of known structure. The 3D structures of these proteins were newly determined by experimental methods, but not yet published or otherwise available to the predictors. The predictions, made from the amino acid sequence alone, thus represent a genuine test of the current performance of threading methods. Only a subset of all the predictions is evaluated here. It corresponds to the 44 predictions submitted for the 11 target proteins seen to adopt known folds. The predictions for the remaining 10 proteins were not analyzed, although weak similarities with known folds may also exist in these proteins. We find that threading methods are capable of identifying the correct fold in many cases, but not reliably enough as yet. Every team predicts correctly a different set of targets, with virtually all targets predicted correctly by at least one team. Also, common folds such as TIM barrels are recognized more readily than folds with only a few known examples. However, quite surprisingly, the quality of the sequence–structure alignments, corresponding to correctly recognized folds, is generally very poor, as judged by comparison with the corresponding 3D structure alignments. Thus, threading can presently not be relied upon to derive a detailed 3D model from the amino acid sequence. This raises a very intriguing question: how is fold recognition achieved? Our analysis suggests that it may be achieved because threading procedures maximize hydrophobic interactions in the protein core, and are reasonably good at recognizing local secondary structure.**

Key words: **protein structure prediction, fold recognition, threading**

INTRODUCTION

One of the most exciting new developments in protein structure prediction has been the approach of identifying whole protein folds from the amino acid sequence. This approach employs techniques for aligning sequences with 3D structures, commonly referred to as *threading,* which are used to select the native fold of a given sequence from a set of alternatives.[1–5] An undeniable catalyst of these developments has been the unprecedented increase in the number of known protein structures. Their ongoing analysis has shown that proteins with very different sequences and functions can adopt very similar folds,[6–8] and, moreover, that known folds fall into a limited set of families.[9–11] This led to the suggestion that the total number of folds accessible to protein sequences is limited,[12,13] which had a particularly important impact on the field of protein structure predictions. It holds the promise of solving the problem of structure prediction, by accumulating sufficient structural data and deriving reliable methods for sequence–structure recognition.

The threading approach embodies methods whose aim is to perform reliable recognition. It has two major components: the actual threading operation, whereby a sequence is optimally aligned onto a given 3D structure, and the criteria for evaluating whether an alignment corresponds to a compatible sequence–structure match.

Both aspects represent very challenging problems to which many groups have recently proposed extremely promising solutions.[1–5] Generating optimal sequence–structure alignments requires handling deletions and insertions, just as in classical sequence alignments. But it is much more complex, because spatial interactions between residues need to be optimized, and not 1D sequence strings. Proposed solutions hence usually rely heavily on dynamic programming algorithms with varying degrees of sophistication.[14,15]

As for the criteria for evaluating sequence structure alignments, they use for the most part poten-

Received for publication June 9, 1995; revision accepted August 31, 1995.

Address reprint requests to Christian Lemer, Unité de Conformation de Macromolécules Biologiques, CP 160/16, P2, Ave. Paul Héger, 1050 Brussels, Belgium.

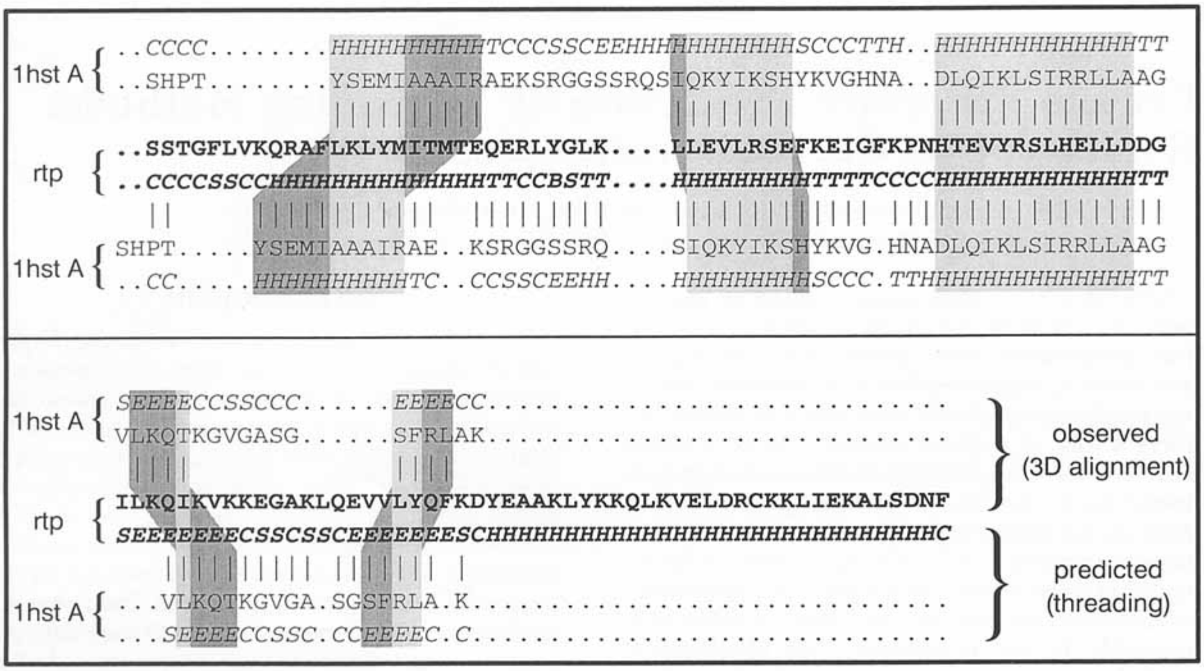

1hst A
..CCCC........HHHHHHHHHHHTCCCSSCEEHHHHHHHHHHHSCCCTTH..HHHHHHHHHHHHHHTT
..SHPT........YSEMIAAAIRAEKSRGGSSRQSIQKYIKSHYKVGHNA..DLQIKLSIRRLLAAG
rtp
..SSTGFLVKQRAFLKLYMITMTEQERLYGLK....LLEVLRSEFKEIGFKPNHTEVYRSLHELLDDG
..CCCCSSCCHHHHHHHHHHHHHHHTTCCBSTT....HHHHHHHHHTTTTCCCCHHHHHHHHHHHHHHTT
1hst A
SHPT.....YSEMIAAAIRAE..KSRGGSSRQ....SIQKYIKSHYKVG.HNADLQIKLSIRRLLAAG
..CC.....HHHHHHHHHHTC..CCSSCEEHH....HHHHHHHHHSCCC.TTHHHHHHHHHHHHHHTT
1hst A
SEEEECCSSCCC......EEEECC...........................
VLKQTKGVGASG......SFRLAK...........................
observed
(3D alignment)
rtp
ILKQIKVKKEGAKLQEVVLYQFKDYEAAKLYKKQLKVELDRCKKLIEKALSDNF
SEEEEEEECSSCSSCEEEEEEESCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHC
predicted
(threading)
1hst A
...VLKQTKGVGA.SGSFRLA.K...........................
...SEEEECCSSC.CCEEEEC.C...........................

Fig. 1. Predicted versus observed alignment for the replication terminator factor rtp(target) and the globular domain of histone H5 (1hst A). The predicted alignment refers to the sequence–structure alignment obtained using threading methods, by the Salzburg group. The observed alignment refers to the 3D alignment of the target and hit structures computed using the MLC version of the program SoFiSt (see Structure Alignment Method). The latter alignment superimposes a total of 39 residues out of the 45 residue of histone H5 with a backbone rms of 1.6 Å (Table III). The amino acid sequence, and the secondary structure assignments obtained by DSSP[19] (shown in italic, above or below each sequence), are given for each protein. The target sequence and secondary structure (bold) are positioned in the middle. Aligned residues are indicated by vertical bars. Segments of secondary structure aligned by SoFiSt (blocks) are shown in dark shading in both alignments. The overlapping portions of the blocks in the two alignments are shaded lightly.

tials derived from the database of known protein structures. These potentials usually contain several contributions, representing residue–residue interactions, or single residue propensities.[1–5] To be effective, they must be able to recognize a cognate alignment from among many incorrect alternatives.

It has so far been difficult to judge how well the different potentials and threading procedures really perform. All of them do equally well in tests where the native sequence–structure combination must be identified from a number of alternatives. But since this number is usually quite limited, and the native match is always perfect, these tests are rather insensitive. A much more challenging task is correctly matching the sequence and 3D structure of two homologous proteins, which tend to differ in length and conformational details as their level of sequence identity decreases. Some of the reported results have been quite encouraging, but overall the performance appears unreliable, particularly at sequence identity levels of 30% or lower, where one would like to see threading outperform classical sequence alignments.

The "Meeting on Critical Assessment of Techniques for Protein Structure Prediction" held in Asilomar, California, in December 1994, offered an unprecedented opportunity to determine the status of current structure prediction methods. These methods fell into three categories: comparative modeling, threading, and ab initio predictions. Using methods in one or more of these categories, 35 laboratories predicted some aspects of the structure of over 30 different proteins. The structures of these "target" proteins were newly determined by X-ray crystallography or NMR spectroscopy, not yet published, and their atomic coordinates were withheld from the predictors until after all the predictions were completed. The predictions were thus performed in a blind fashion, with the expected results not known beforehand.

This paper evaluates the predictions in the threading category. It presents a critical overview of the results obtained by 9 different groups, using a variety of threading approaches, on a total of 21 target proteins, representing 6 different folds.

ANALYSIS OF THE TARGET STRUCTURES

The Target Protein Structures

Information on the 21 protein chains, which were the target structures for this contest, is summarized in Table I. Structures which were made available, but for which no predictions were submitted, were not considered. In the remainder of this paper, and unless specified otherwise, the target proteins will be referred to by the shorthand notation given in column 1 of Tables Ia and Ib.

To qualify for the threading contest, the 21 target structures were selected so as to display no significant sequence identity with any protein of known structure. Since the predictions were blind, threading operations were performed on sequences irrespective of whether they correspond to a known fold. Here we assessed only predictions for target proteins which do correspond to such folds. The first step of the assessment procedure was therefore to determine which of the target structures adopt known folds, and what these folds are.

The Representative Set of Known Protein Folds

To determine the known fold, if any, that a given target structure resembles, the structure was compared to a representative set of folds from the Brookhaven Protein Data Bank.[16] This set consisted of 203 structures representing the different protein folds known to a relatively recent date.[10]

Structure Alignment Method

The structure alignments were performed using the automatic procedure SoFiSt, recently developed in our laboratory.[17,18] In this procedure, the quality of the alignment is evaluated using, as sole measure, the root mean square deviation (rms) between backbone atoms (N, Cα, C, and O) superimposed. The algorithm comprises three main steps: (1) identifying segments of secondary structure (α-helix and β-strands), defined by DSSP,[19] which have similar conformations in both 3D structures; (2) applying a generalized hierarchic clustering procedure to find collections of these segments which yield optimal global structure alignments; and (3) extending the alignments to include residues outside the initial segments.

Two different versions of the algorithm were used. A very accurate version where the clustering step is performed using a constrained Multiple Linkage Clustering (MLC) algorithm, yielding optimized solutions, and a somewhat less accurate, but faster, version which uses an Intermediate Linkage Clustering (ILC). This faster version was used to scan the database of representative proteins with each target structure to identify the most resembled fold, and the accurate version was used to obtain an optimal alignment of the target structure with this fold. Note that we were able to verify that the alignments produced by the MLC and ILC versions showed only slight differences.

Identifying Target Structures That Correspond to Known Folds

Each target structure was systematically aligned to the 203 representative folds in our database, using the automatic ILC procedure. This procedure was applied, requiring that a given pairwise global alignment displays an rms of 3 Å or less, and that the smallest protein in the aligned pair has at least 50% of the residues aligned in its secondary structures. This usually produced several acceptable solutions wherein the target structure was superimposed on different representative folds. The "best" solutions obtained for each of the target structures are listed in the second to last column of Tables Ia and Ib.

The decision on whether a target structure resembles one of the representative folds was not easy to make. Such a decision should ideally be based on objective measures which can distinguish random alignments from significant ones. With such measures still under development at present, the judgment was made on the basis of several criteria applied to the "best" solutions obtained by SoFiSt. These were (1) the number of aligned secondary structure segments, (2) the rms of the alignment, and (3) visual inspection of the superimposed structures.

Out of the 21 target structures, the 12 structures listed in Table Ia were judged to have structural homologs in our dataset. This includes the modeled structure, Mystery. The remaining 9 structures, listed in Table Ib, were considered to represent new folds whose structural similarity to known folds was judged not to be significant. Unless specified otherwise, identified known folds are referred to by their Protein Databank codes throughout this study.

Visual inspection was particularly useful in some cases. For example, the 2 target proteins kau A and ppdk 3 (Table Ia) were seen to represent known folds, even though only a portion of their 3D structure was found to resemble a fold in our database. For kau A, this portion was found to contain part of a TIM barrel, whereas for ppdk 3 it contained a composite fold (α + β sandwich and β meander).

The Appendix shows ribbon drawings of the 11 experimentally determined target structures identified as having known folds (Table Ia), and the structure most similar to each target, identified from among the 203 proteins of the dataset by the structure alignment procedure SoFiSt.

INFORMATION MADE AVAILABLE TO PREDICTORS

The teams participating in the contest were expected to predict the structure of the target protein, or proteins, of their choice from the amino acid sequence alone, using threading methods. In all cases information on the amino acid sequence of each protein was provided. However, in some instances the

TABLE Ia. Target Proteins Corresponding to Known Folds*

Code	Protein name	Number of residues	Secondary structure	Fold	Structure alignment	Solved by
staufen3	Domain 3 of staufen *Drosophila*	68	2α 3β	α + β sandwich	1pda 2α 3β (2.2 Å)	M. Bycroft
prosub	Propeptide from subtilisin BPN^{1}	73	2α 4β	α + β sandwich	2nck L 2α 4β (1.9 Å)	T. Gallagher, P. Bryan et al.
kau B	Urease (chain B) *Klebisella aerogenes*	101	1α 8β	β-sandwich	2bpa L 0α 6β (2.8 Å) 4sbv A 1α 5β (2.4 Å)	E. Jabri, A. Karplus
rtp	Replication terminator protein *Bacillus subtilis*	118	4α 2β	Histone-like	1hst A 3α 2β (1.6 Å)	D. Bussiere, S. White
synapto	Synaptotagmin (first C2 domain)	126	1α 8β	β-sandwich	1cob A 0α 6β (2.5 Å)	S. Sprang
ppdk 3	Pyruvate phosphate dikinase (domain 3: 370–520)	151	4α 8β	α/β	1aco 3α 6β (2.4 Å)	O. Herzberg
pcna	DNA polymerase processivity factor *yeast*	261	4α 18β	DNA clamp	2pol A 4α 16β (2.2 Å)	T. Krishna, X.-P. Kong et al.
xylanase	Xylanase (catalytic core) *Pseudomonas fluoresences*	345	9α 13β	TIM barrel	1btc 8α 7β (2.5 Å)	G.W. Harris, J. Jenkins et al.
ppdk 4	Pyruvate phosphate dikinase (domain 4: 520–874)	360	18α 8β	TIM barrel	5tim A 7α 6β (2.3 Å)	O. Herzberg
pbdg	6-Phospho-β-D-galactosidase *Lactococcus lactis*	455	15α 16β	TIM barrel	1nar 7α 8β (3.0 Å)	C. Wiesmann
kau A	Urease (chain A) *Klebsiella aerogenes*	566	16α 27β	Central domain: Part of TIM barrel	5tim A 5α 8β (2.8 Å)	E. Jabri, A. Karplus
mystery*	Model protein	182	8α 8β	Modeled to be a TIM barrel	1gox 8α 8β (2.9 Å)	—

*The shorthand notation used to designate the target proteins throughout this paper is given in column 1. The protein names are given in column 2 and their number of residues in column 3. The number of α-helices and β-strands in each protein, as assigned by DSSP,[19] is given in column 4 and the protein fold defined according to CATH[10] is in column 5. Column 6 gives the Protein Data Bank code of the most similar structure among the set of 203 known folds (see Methods), as identified by the 3D structure alignment procedure SoFiSt.[17,18] The protein names corresponding to the PDB codes are 1pda (porphobilinogen deaminase), 2nckl (nucleoside diphosphate kinase), 2bpa (bacteriophage ϕX174 capsid protein), 4sbv (southern bean mosaic virus coat protein), 1hst (histone H5), 1cob (superoxide dismutase), 1aco (aconitase), 2pol (polymerase III), 1btc (β-amylase), 5tim (triose phosphate isomerase), 1nar (narbonin), 1gox (glycolate oxidase). Column 6 also lists data on the 3D alignment, which are the number of α-helices and β-strands superimposed with the target, and the corresponding backbone rms deviation. The dataset scan is performed with the faster ILC version of SoFiSt. But the described alignments of the top hits are obtained with the MLC version, which gives optimal solutions with respect to the rms of superimposed backbone atoms. The rightmost column lists the authors of the experimental determination (by X-ray crystallography or NMR) of the target 3D structure.

TABLE Ib. Target Proteins Corresponding to New Folds*

Code	Protein name	Number of residues	Secondary structure	Structure alignment	Solved by
L14	Prokaryotic ribosomal protein *Bacillus stearothermophilus*	112	1α 7β	1sry A 1α 4β (2.77 Å)	S. White, C. Davies
bhted	β-Hydroxydecanoly thiol ester dehydrase	171	3α 7β	1fus 1α 3β (2.16 Å)	M. Leesong, J. Smith
bphc	Biphenyl$_{2,3}$-diol$_{1,2}$-dioxygenase *Pseudomolas* sp. strain LB 400	188	3α 9β	2sn3 1α 3β (2.62 Å)	J. Bolin
ce-1	Chymotrypsin/elastase inhibitor-1	61	0α 4β	1hsb A 0α 4β (1.74 Å)	K. Huang, M.N. James
chmut	Chorismate mutase *Escherishia coli*	95	3α 0β	1pgd 3α 0β (2.67 Å)	J. Clardy
kau G	Urease (chain G) *Klebsiella aerogenes*	100	4α 2β	1ezm 3α 0β (2.44 Å)	E. Jabri, A. Karplus
ppdk 1	Pyruvate phosphate dikinase (domain 1: 1–125)	250	10α 7β	3rub S 2α 3β (2.51 Å)	O. Herzberg
ppdk 2	Pyruvate phosphate dikinase (domain 2: 250–370)	121	4α 5β	1hsb A 2α 3β (2.88 Å)	O. Herzberg
smanucecs	Extracellular endonuclease *Serratia marcescens*	250	6α 8β	1ten 0α 5β (2.62 Å)	M. Miller, J. Tanner et al.

*These represent targets for which 3D structure alignment procedures could not identify a similar fold in our dataset of known structures or found structural similarities which we judged as not significant (e.g., ce-1, chmut). The columns are the same as in Table Ia, except for the missing column 5, which gives the fold class. The protein names corresponding to the PDB codes given in column 5 are 1sry (seryl-trna synthetase), 1fus (robnuclease f1), 2sn3 (scorpion neurotoxin), 1hsb (class I histocompatibility antigen) 1, pgd (6-phosphogluconate dehydrogenase), 1ezm elastase), 3rub (ribulose 1,5-biphosphate carboxylase), 1ten (tenascin).

TABLE II. The Predictor Teams*

Code	Predictors	Number of predicted targets	
		Submitted	Assessed
Baltimore	N. Clarke (acknowledging J. Bowie)	2	1
Oxford	R.R. Copley, C.D. Livingstone, R.B. Russell, G.J. Barton	3	3
Scripps	A. Godzik	12	6
Cambridge	T.J.P. Hubbard, J. Park	6	4
London	D.T. Jones, R.T. Miller, J.M. Thornton	17	9
NIH	T. Madej, S.H. Bryant, J.-F. Gibrat	9	7
Osaka	Y. Matsuo, K. Nishikawa	19	10
Salzburg	M.J. Sippl, H. Floeckner, M. Braxenthaler, P. Lackner, M. Jaritz, M. Ortner	15	8
EMBL	M. Willmanns, D. Eisenberg	3	1

*Column 1 lists the code by which the different teams are referred to in this paper. Column 2 lists the members of each team, who contributed to the predictions. Column 3 lists the number of predictions submitted by each team. The submitted predictions include all the threading predictions submitted by each team. Only a subset of these predictions, those made for the targets which correspond to known folds, was assessed.

predictors were referred to the SWISSPROT or PIR databanks to get this information, while in others the amino acid sequence was directly provided. The former situation made the assessment task somewhat more difficult, due to occasional differences between the sequence used in the threading and that of the target structure.

Other information made available to predictors included the protein name, its source, known sequence homologies, and the author names. In some, rare cases, the authors provided references to papers which suggested possible folds for the target sequence.

ASSESSMENT PROCEDURE

Data Made Available to the Assessors

The following information was made available to the authors of the present paper for the purpose of assessing the prediction results. For each predicted target protein, the predictors provided (1) the library of known structures used in the threading tests, (2) the 10 best hits obtained using this library, (3) the scores, preferably the Z-scores, for the 10 best hits, and (4) the sequence-structure alignment for the best hit.

In cases where the structure of loops was also predicted, the full atomic coordinates of the predicted protein was also provided. Assessment of these data was not made, however, because the prediction accuracy of the core regions was, in general, insufficient to justify a detailed analysis of the loops. In addition to the above information, predictors also produced a succinct description of their results, which was sometimes accompanied by a short discussion.

The Two-Level Assessment

Only the predictions on the 11 target structures corresponding to known folds (Table Ia) were assessed. The assessment was done at two distinct levels: the fold and the sequence–structure alignment. First we determined if the correct fold was among the 10 top hits of each predictor. Then the quality of the sequence–structure alignment was analyzed in the thread corresponding to the top hit. In assessing fold recognition we chose to examine the 10 top hits instead of just the first one, to allow for more flexibility in the evaluation.

Are predicted folds similar to the target structure?

To determine which of the top 10 hits, if any, corresponds to the correct fold, the structures of these hits were successively aligned to the corresponding target structure, using the ILC version of the SoFiSt structure alignment procedure. A fold was considered as correct when a significant fraction of the secondary structure elements of the predicted structure was aligned to the target, with an rms $\leq$ 3 Å.

To judge whether a fold was correct or not, we also consulted the classifications of protein folds, CATH (following Orengo et al.[10]) and SCOP.[20] The top 10 hits were assigned to known fold families, and possible similarities between them were identified.

Evaluating the sequence–structure alignments

The quality of the predicted sequence–structure alignment was evaluated for all top hits corresponding to correctly identified folds, and for which we had the alignment data. This was done by measuring, for each target–hit pair, the correspondence between the 3D arrangements of residues in secondary structures deduced from the predicted alignment and from the 3D structure alignment generated by the MLC procedure in SoFiSt.

To gain insight into how threading achieves fold recognition, and which factors influence the sequence–structure alignments, we compared the per-residue secondary structure assignments in the predicted versus observed alignments. In addition, we computed the content of hydrophobic residues in the protein core in the predicted alignments, and compared them to those observed in the cores of the native hit proteins. Note that when evaluating the predicted spatial arrangements, only portions of the hit structures, corresponding to the parts aligned by SoFiSt, were taken into account. In comparing secondary structure assignments, or core composition, the predicted alignment, which usually extended over the whole protein, was considered in its entirety.

Spatial arrangement. To evaluate the 3D arrangement, a rather crude measure was used. We examined to what extent threading matches the same secondary structure elements of the target protein to the hit, as SoFiSt does.

Secondary structure assignments were derived for the target and the top hit structures in each prediction using DSSP.[19] The secondary structure segments of a given target–hit pair aligned by SoFiSt before the extension step (see Structure Alignment Method) were defined as *blocks.* A *block* in the hit protein was considered as correctly aligned by threading when at least one of its residues overlapped the same block in the target, to which it was aligned by SoFiSt (see Fig. 1).

As a quality measure of the 3D arrangement, we used the ratio of the number of correctly aligned blocks by threading over the total number of blocks aligned by SoFiSt. TIM-barrels represented a special case. Due to their high symmetry, circular permutations of the $\alpha\beta$ units in one structure relative to the other have little effect on the quality of the structure alignments. To take this into account, circularly permuted blocks were not penalized in the predicted alignments of these folds.

To evaluate the overlap between a correctly aligned threaded block and its equivalent aligned by

SoFiSt, we computed the number of residues by which this block was shifted along the sequence relative to the SoFiSt equivalent (see legened of Fig. 1 for detail), and calculated the average shift of the correctly aligned blocks in a given thread.

The MLC version of SoFiSt, used here, usually generates several optimal structure–structure alignments. These alignments differ somewhat by their rms deviations and the number of aligned residues.[17,18] The number of correctly aligned *blocks* and their average shift will depend on which of these alignments is used as reference. Given this choice, we systematically favored the predictors, by taking as reference the SoFiSt alignment closest to the threading solution.

Secondary structures. From the predicted sequence–structure alignment, we assigned the secondary structure of the target. This predicted assignment was compared to the corresponding DSSP assignment[19] derived from the atomic coordinates of the target. To measure the agreement we computed the usual 3-state identity score Q_3. This score was shown not to exceed 88% on the average, for homologous proteins with essentially identical 3D structures.[21] This top limit should be lower when dealing with pairs of homologous structures from unrelated proteins. In addition, we computed the percentage of misprediction, defined as predictions of the wrong secondary structure type: an α-helical residue predicted to be in a β-strand and vice versa.

Protein core composition. From the predicted alignment of the target sequence onto the hit 3D structure, we computed the fraction of the residues positioned in the protein core that is hydrophobic. This predicted fraction was compared to that observed in the native hit protein. The protein core was defined as the residues in secondary structures which bury 70% or more of their solvent accessible surface area. The surface area calculations were performed on the atomic coordinates of the hit protein, using the analytic algorithm SurVol.[22] The secondary structure assignments were obtained by DSSP. The following residues were defined as hydrophobic: Ala, Val, Ile, Leu, Met, Phe, Pro, Tyr, Trp.

RESULTS AND DISCUSSION

Nine teams submitted threading predictions for the target proteins of their choice, representing a total of 86 predictions on 21 different target proteins (Table II). The number of targets predicted by each group was highly variable, ranging from 2 by the Baltimore group to 19 by the Osaka group. Our evaluation focused only on the 44 predictions made for the 11 target structures, judged by us as having known folds (Table Ia). As a matter of consistency, we established that the relevant fold, or its close homolog, was included in the library of known structures used by each predictor.

Assessing the threading predictions has been very stimulating, but also an extremely challenging undertaking. Because threading is a relatively recent approach, there are as yet no well established criteria for measuring its performance. Such criteria require ways of comparing sequence–structure matches with structure–structure matches. The problem is particularly challenging when one deals with proteins that are not obviously related, owing in part to the inherent difficulty in defining how similar two structures really are in these cases.[23] More time and analysis of additional data, not available to this assessment, would have helped to deal with these issues better than we have. One should therefore regard our evaluation as a rough first attempt at presenting a synthetic picture of the current performance of threading methods.

This section comprises three main parts, aiming at answering the following questions: Can threading predict the correct protein fold from the amino acid sequence? How good are the predicted sequence–structure alignments? Are some threading methods more effective than others?

Can Threading Predict the Correct Protein Fold?

The known structures, which our structure–structure alignment procedures identify as most similar to each target protein, are listed in the second to last column of Table Ia. This column also describes their aligned secondary structures, and the rms value of the alignment. We see that these structures represent a total of 6 folds. A significant number of these corresponds to very common folds. More than half has the TIM barrel fold, and another 4 have folds in the α + β sandwich and β-sandwich families. None of the targets is fully helical.

Table III lists the folds predicted for each target by the predictor teams. These folds do not always correspond to the top hit of each team, but to the first structure in the list of the 10 top hits, which is similar to the target, as judged by our structure alignment procedures.

Different measures of the fold prediction performance

The results in Table III can be viewed in a number of different ways, each leading to quite different conclusions.

A conservative view consists in considering a threading prediction as successful only when a fold similar to the target structure is the top hit, thus, when it corresponds to the thread with highest score, or equivalently, with lowest energy. Judging by this definition, the prediction results are rather uneven, and some groups seem to perform much better than others.

To score this performance, one could use the ratio of the correct over the total number of predictions made by each group. This measure may, however,

TABLE III. Fold Prediction Results*

	Osaka	London	Salzburg	NIH	Scripps	EMBL	Baltimore	Cambridge	Oxford
staufen3	5 2yhx *1α 3β (1.8)*	>10	8 2sni i *2α 2β (2.1)*	9 1tpb a *1α 3β (2.1)*	**1 1npc** ***1α 3β (2.5)***	▨	▨	**1 2cmd** ***2α 2β (2.7)***	▨
prosub	>10	▨	**1 2fxb** ***2α 3β (2.9)***	2 1raa b *2α 4β (2.3)*	▨	▨	▨	▨	▨
kau B	**1 2mcm** ***6β (2.2)***	**1 2rhe** ***4β (2.0)***	▨	▨	▨	▨	▨	2 3cms *5β (2.7)*	1 2fbj l *5β (2.7)*
rtp	>10	>10	**1 1hst a** ***3α 2β (1.6)***	10 1hst a *3α 2β (1.6)*	▨	▨	▨	▨	▨
synapto	6 1cd8 *7β (2.4)*	**1 1cd8** ***7β (2.4)***	>10	▨	8 7pcy *7β (2.3)*	▨	**1 2tbv** ***6β (2.2)***	4 1mcp h *6β (1.9)*	>10
ppdk 3	3 1add *3α 5β (2.4)*	6 2dnj a *2α 5β (2.8)*	▨	**1 1etu** ***3α 4β (3.0)***	4 1npx *2α 5β (2.8)*	▨	▨	▨	▨
pcna	▨	▨	>10	▨	▨	▨	▨	▨	▨
xylanase	▨	>2	**1 1tim b** ***7α 8β (2.8)***	▨	▨	▨	▨	1 1xla a *6α 7β (2.9)*	▨
ppdk 4	**1 1pii** ***8α 8β (2.5)***	**1 1gox** ***8α 8β (2.4)***	▨	8 1btc *8α 7β (2.8)*	3 8rub l *8α 7β (2.8)*	>1	▨	▨	▨
pbdg	3 1pii *5α 8β (2.7)*	**1 1add** ***5α 7β (2.7)***	10 1nar *7α 8β (2.9)*	2 2tmd a *8α 7β (3.0)*	>10	▨	▨	▨	▨
kau A	4 1cdg *5α 7β (3.0)*	**1 1pii** ***4α 6β (2.5)***	▨	▨	▨	▨	▨	▨	▨
mystery	2 1chr a *7α 7β (2.7)*		5 1pii *7α 7β (2.3)*	10 1pii *7α 7β (2.3)*	4 trea *7α 8β (2.8)*				1 5rub a *8α 8β (2.6)*

*Each row refers to a target protein. Its shorthand notation, defined in Table Ia, is given in column 1. Each column refers to a predictor team, denoted by their code, given in Table II. Each box contains information on a given predition experiment. It gives the position in the score ranking of the first correct hit, followed by the PDB code of this hit. Underneath are the number of α-helices and β-strands of the hit, which are aligned with the target by the optimal 3D alignments obtained with the MLC version of SoFiSt. The rms (in Å) of the corresponding alignment is given in parentheses. Box shading indicates the position of the correct fold in the score ranking. White boxes indicate that the correct fold was the top hit; light gray boxes that it was among the 10 best hits. Dark gray boxes indicate that the correct fold did not appear among the hits provided by the predictors; ">1," ">2," and ">10" indicate that the correct fold was not among the 1st, 2nd, and 10 best hits, respectively. Certain preditors provided us with either the *best* hit, or the top 2 *best* hits only. Hashed boxes indicate that no prediction was entered for the particular target by the corresponding predictor. The entire row for the target mystery is shaded to indicate that the predictions of this modelled protein were not assessed.

not be very meaningful, because some groups, such as Osaka, London, and Salzburg, contributed 7–9 predictions, whereas other groups contributed far fewer predictions.

If we nevertheless consider such a ratio, two groups, London and Salzburg, clearly stand out, with a ratio of 5/9, and 3/7, respectively. Osaka, Scripps, and NIH do less well, with ratios of 2/9, 1/5, and 1/6, respectively.

The Oxford and Cambridge groups score a 50% success rate on a total of 2 and 4 predictions, respectively. But their prediction methods are sufficiently different from the other threading approaches analyzed here (see Summary of the Prediction Methods) not to warrant direct comparison. EMBL and Baltimore contributed only 1 prediction each, and could therefore not be evaluated.

A more lenient view consists in considering a threading prediction as successful if a fold similar to the target is among the top 10 hits. An objective basis for or against such a view may, in principle, be obtained from evaluating the odds of positioning by chance a fold similar to the target among the top 10 hits. These odds depend on the fraction of the motifs in the library which corresponds to folds similar to the target. This fraction may differ for each prediction experiment and for different authors. More importantly, it depends on the stringency with which fold similarity is defined. Considering our set of 203 representative folds as typical, and taking the urease B chain target (kau B) as an example, we see that our structure similarity criteria (see Methods) identify 8 folds (4%) as being similar to kau B. If we also count the structural neighbors of these folds, as defined by the CATH classification, than our set contains 41 folds (~20%) resembling kau B. Thus the odds of identifying the correct fold by chance is 1 in 25, and 1 in 5 for the first and second similarity criterion, respectively, which in both cases is quite significant. A similar situation is encountered with synapto, ppdk3, and prosub. These targets have, respectively, 7, 16, and 5 similar folds in our dataset according to the structure alignment results, and over 24, 12, and 20 similar folds, respectively, according to looser criteria following CATH. Identifying a correct fold among the top 10 hits for these 4 targets is thus not statistically significant. It could still be significant, however, for the remaining 6 targets, for which our structure alignments procedure identifies on the order of only 1–3 similar folds in a total of 203 structures.

A different argument in favor of the lenient view comes from the observation that in all the analyzed predictions for which scores were provided, the top 10 hits all have very similar scores. As far as we could determine, scores of correct predictions were thus not consistently higher than for incorrect ones. This suggests that, in general, the exact position of the correct fold in the list of top hits is probably somewhat arbitrary.

Using the lenient view then, a quite different picture can be drawn of the threading results, and of the performance of each group. Among the predictors who use conventional threading, the NIH group now performs best. A fold similar to the target always appears among their top 10 hits. The performance of Scripps and Osaka is also significantly improved, with scores of 4/5 and 7/9, respectively. Interestingly, the London performance remains essentially unchanged. They identify the correct fold as their top hits or miss it altogether. On the other hand, the Salzburg performance improves significantly relatively to the conservative evaluation, but since the performance of other groups improves more dramatically, it ceases to stand out. Lastly, the performance of Cambridge is significantly improved. That of Oxford could not be, as this team submitted only one entry per trial.

Fold recognition is not all-or-none

Detailed inspection of Table III suggests that the degree of similarity between the predicted and the target structures may vary significantly. The structural similarity is nearly perfect between the target rtp, the replication terminator protein from *B. subtilis*, and its predicted fold, the globular domain of histone H5 (1hst A). Three α-helices and 2 β-strands of the structures align with a 1.6 Å rms. Both Salzburg and NIH identified this perfect match. But Salzburg had the histone as the top hit, whereas NIH had it at position 10.

In predictions made for other proteins, the similarity of the highest scoring folds to the corresponding targets was less obvious and often seemed borderline. For example, our 3D structure alignments show that 2rhe, the fold identified by London for the target urease B chain, kau-B, has no more than 4β-strands well aligned with the latter (2.0 Å rms). Since kau-B comprises a total of 8 strands and 1 helix, less than half of its structure resembles the identified fold. Counting this and other similar cases as correct fold predictions was done to help the predictors. It furthermore takes into consideration the fundamental problem of making a judgment about structural similarity in proteins with no obvious evolutionary relationship.[23] The same problem was also encountered in identifying the known folds associated with each target.

Are certain folds easier to predict?

The structure libraries used by some predictors contained several copies of the common folds, also termed superfolds,[13] and often only a single copy of the less common ones. It is therefore reasonable to assume that this would lead to better prediction scores for targets that adopt common folds, due either to higher odds of identifying the correct fold by chance, as discussed above, or to biases in the potentials and scoring schemes.

We see that all the targets which correspond to common folds had their fold usually predicted correctly. On the other hand, rtp, the small protein with the less common histone-like fold, and pcna, which resembles only one known structure, the β-subunit of *E. coli* DNA polymerase III (2pol), are the least well predicted targets. Of the 4 predictions submitted for rtp, by Osaka, London, Salzburg, and NIH (Table III), 2 are mispredictions (the correct fold is not among the 10 top hits), representing a 50% failure rate.

The case of pcna is somewhat special, and possibly revealing. It was the only target protein larger than 180 residues which was not a TIM barrel. Only the Salzburg team submitted a prediction for this protein. Other groups were unable to add the known *E. coli* DNA polymerase structure to their fold libraries in time for the prediction, as this structure was released during the latter part of 1994. The Salzburg result was nevertheless incorrect. They predicted pcna to be a TIM barrel, even though the DNA polymerase β-subunit was present in their structure library. In fact, the structural homology between pcna and the DNA polymerase clamp was predicted by the crystallographers,[24,25] notwithstanding the insignificant sequence identity.

A different manifestation of the TIM barrel bias is exemplified by the prediction results for mystery. This target is a model TIM barrel, de novo designed by the first EMBO protein design course and subsequently amended by Chris Sander and Gerrit Vriend (Chris Sander, personal communication). However, the corresponding protein was produced, and shown by CD and NMR to have almost no β-sheet (Steve Emery, personal communication), and thus not to adopt the TIM fold.

We see that all 5 predictions for this model protein identify a TIM barrel among the top 10 hits (bottom of Table III). Interestingly, the Oxford group, whose methods rely on consensus secondary structure predictions of homologous sequences, identify a TIM barrel as the top hit, whereas predictors using the more conventional threading methods identify a

TABLE IV. Quality of the Predicted Sequence–Structure Alignments*

			Predicted versus observed structure alignment		Predicted versus observed secondary structure		Hydrophobics in core of hit
Predictor	Target	Best hit	Number of blocks predicted/observed	<shift>	Q_3 (%)	Misprediction (%)	In sequence of target/hit (%)
Osaka	kau B	2mcm	0/4	♦	56	0	58/61
	ppdk4	1pii	13/16	—	47	26	55/68
London	kau B	2rhe	0/3	♦	59	5	75/64
	synapto	1cd8	2/6	1.5	70	0	80/69
	ppdk4	1gox	8/16	—	43	24	57/69
	pbdg	1add	6/15	—	50	28	55/60
	kau A	1pii	8/10	—	45	28	58/68
Salzburg	prosub	2fxb	3/4	0.7	64	21	87/67
	rtp	1hst A	5/5	2.2	84	0	69/69
	xylanase	1tim B	8/16	—	61	25	62/66
NIH	ppdk3	1etu	2/6	2.5	31	28	58/65
Cambridge	staufen3	2cmd	0/3	♦	57	17	37/63
	xylanase	1xla A	10/16	—	61	10	55/62
Oxford	kau B	2fbj L	2/4	1.5	58	0	75/57

*Data in this table concern only the sequence–structure alignments obtained for the top hits corresponding to correct folds. These hits are those listed in the white boxes of Table III. Columns 1 (*predictor*), 2 (*target*), and 3 (*best hit*) list the predictor team, the target protein, and the top hit of each predictor, respectively. Columns 4 and 5 (*predicted versus observed structure alignment*) summarize the results of the comparisons between the threading alignments (predicted) and the (observed) 3D structure alignments obtained using the optimal MLC version of SoFiSt. Column 4 (*number of blocks predicted/observed*) lists the ratio of correctly matched secondary structure segments (blocks) in the threading alignment over the total number of superimposed blocks in the 3D alignment. Column 5 (<*shift*>) lists the average number of residues by which the correctly aligned blocks are shifted along the sequence, relative to the blocks aligned by SoFiSt. ♦ indicates that the shift computation was not applicable, as none of the blocks was aligned correctly.— indicates that the shifts could not be computed because the target is a TIM barrel. In such cases predicted alignments corresponding to circular permutations of the βα motifs were considered as correct. But it was not possible to compute all corresponding 3D structure alignments by SoFiSt. Column 6 and 7 (*predicted versus observed secondary structure*) compare the secondary structures deduced from the predicted alignment of the target sequence on the hit structure, with the observed secondary structure of the target. DSSP[19] was used to assign the secondary structures in both the hit and target proteins. Column 6 (Q_3) gives the per-residue 3 state (α, β, and other) identity score Q_3. Column 7 (*misprediction*) gives the percentage of mispredictions: where an α-helical residue was predicted to be in a β-strand and vice versa. The numbers in the rightmost column (*hydrophobics in core of hit*) represent the percentage of the residues buried in the core which are hydrophobic, when the target sequence is aligned on the structure of the hit (numbers on the left), and when the native hit sequence is aligned (numbers on the right). The protein core was defined as the residues in secondary structures which bury 70% or more of their accessible surface area to solvent. The sequence–structure alignments of the Scripps and Baltimore groups were not assessed. The former did not provide alignments, whereas the latter provided alignments for an erroneous prediction selected by visual inspection (see text).

TIM fold further down their hit list. This suggests that the designed TIM barrel reflects better the secondary structure prediction rules, than the rules that may govern tertiary interactions.

Quality of the Sequence–Structure Alignments

Are residues in secondary structure elements correctly positioned in space?

An important aim of fold predictions by threading is to derive a valid description of the 3D structure of a given sequence in terms of its atomic coordinates. This would require the sequence of the target to be aligned onto the homologous fold so as to place its residues in a spatial arrangement similar to its native structure.

To evaluate the degree to which this was achieved in the threading predictions, we compared the predicted sequence–structure alignments with the observed structure–structure alignments obtained by the MLC version of SoFiSt. This was done only for the predictions in which the correct fold was identified as the top hit, and when alignment data were made available. The alignment analysis is summarized in Table IV.

A rough, yet revealing, criterion whereby the correspondence between the predicted and observed alignments can be evaluated, is the ratio of correctly matched secondary structure segments (blocks) over the total number of superimposed secondary structure segments in the observed alignment. Even by this rather crude measure, which considers a block as correctly matched when it overlaps by a single residue the block aligned by SoFiSt, very few predictions achieve correct alignments.

By far the best result obtained in the entire contest is the rtp/histone (1hst A) alignment predicted by Salzburg. This alignment matches correctly all the 5 blocks aligned by SoFiSt (Table IV), and could hence be considered as perfect. However, though the average shift of these blocks is 2.2 residues, which may seem low, detailed inspection reveals that significant local shifts of some blocks do occur.

In particular, the first 10-residue helix of 1hst A is shifted by 5 residues (Fig. 1). As a result, the overlapping segment between the predicted and observed alignments extends over half the helix only. Furthermore, this 5-residue shift corresponds to a rotation of ~90° about the helix axis. This affects the helix orientation relative to the other helices of the histone motif, whose alignment is essentially perfect (helix 2 is shifted by only one residue; helix 3, not at all) and could modify the packing of the hydrophobic core.

The 2β-strands, which form a hairpin (see Fig. 2, d1 and d2, in the Appendix), are also shifted by 3 and 2 residues, respectively. These strands, comprising 4 residues each, are thus significantly shifted in comparison to their length. The 3-residue shift in the first strand introduces a potentially destabilizing phase shift in the orientation of hydrophobic side chains on the hairpin surface.

In addition, we observed that all the shifts described above occur when the equivalent secondary structures in the target and the hit differ in length. They happen because the threading and the 3D structure alignments match different portions of the longer secondary structures of the target with the shorter secondary structures of the hit (Fig. 1). Interestingly, therefore, though these shifts may alter tertiary interactions, they completely preserve the local secondary structure.

In all the other predictions, only a subset of the secondary structure segments superimposed by the 3D structure alignments was correctly aligned by threading. There were even 3 predictions, those for kau B by Osaka and London, and the staufen3 prediction by Cambridge, which had none of the secondary structures segments correctly aligned.

This rather puzzling result immediately raises a very intriguing question: how could the correct folds be recognized when only a few, or none, of the secondary structure elements were correctly positioned in space?

To gain insight into the origins of this behavior, and to follow up on the observation that the shifts of correctly aligned blocks often tend to preserve local secondary structure, 2 additional aspects of the alignments were analyzed. We compared on a per-residue basis the secondary structure assignments deduced from the predicted alignment to the corresponding quantities in the target structure. In addition, the predicted alignment was used to compute the residue composition in the core of the hit structure. This composition was compared with that observed in the native hit protein. This was done for all the target–hit pairs of Table IV.

Does the alignment of local secondary structure play a role in fold recognition?

The results of the per-residue 3-state identity scores for secondary structure (Q_3) are given in column 6 of Table IV. We see that the highest Q_3 score is obtained for the excellent rtp/1hst A alignment predicted by Salzburg. Its value (84%) is very close to the upper limit expected for a pair of homologous proteins with nearly identical 3D structures.[21] On the other hand, the ppdk 3/1etu alignment predicted by NIH has the lowest score of 31%, a value corresponding to random predictions.

In a number of other cases, however, quite reasonable Q_3 scores are achieved for target–hit pairs with very poorly predicted spatial alignments. In particular, the very poor alignments of kau B/2mcm (Osaka), kau B/2hre (London), staufen3/2cmd (Cambridge), and synapto/1cd8 (London), in which none, or only a few of the secondary structure segments are correctly aligned in space, have Q_3 scores of 56, 59, 57, and 70%, respectively. The low misprediction score in these alignments is mostly due to the fact that, except for the staufen3/2cmd pair, these proteins contain essentially β-strands.

Quite low Q_3 scores (43–47%) and a high level of misprediction (24–28%) are observed for the predicted alignments of all TIM barrel targets except, perhaps, xylanase. This seems at odds with the relatively reasonable spatial alignments predicted for these proteins, but can be explained by the fact that we allow for circular permutations in these alignments. It is also clear that mispredictions are expected to occur much more readily in α/β proteins, as even small shifts along the sequence can result in α↔β substitutions.

Interestingly, the Q_3 scores achieved by Oxford and Cambridge are of the same order as those obtained by other groups. This is more surprising for Oxford, since their method relies explicitly on secondary structure predictions of the target protein. Cambridge, on the other hand, uses secondary structure overlap criteria only to rerank the alignments, after the latter were ranked by their alignment score.

Finally, it is noteworthy that the results are not significantly altered when the residues at the beginning and end of each secondary structure element are not counted in computing the Q_3 scores. This indicates that the errors are not confined to the boundaries of the secondary structure elements, as it is often the case for the inconsistencies between secondary structure assignments in pairs of homologous proteins.[26]

We very tentatively deduce from this analysis that seemingly serious imperfections in the spatial alignments can be counterbalanced by contributions from local interactions along the chain, which are taken into account in the effective potentials used in most threading approaches. Such contributions can be computed from backbone conformational preferences, or by compiling residue pair interactions separately for different distances along the sequence,

TABLE V. Summary of the Database-Derived Potentials Used by the Predictors*

		Osaka	London	Salzburg	NIH	Scripps	EMBL	Baltimore
Potential terms	**residue pair interactions**	○	○	○	○	○		
	contact / non contact					✓		
	distance – dependent	✓	✓	✓	✓			
	atoms considered	C_β	C_β/C_β,N,O	C_β	$C_{\alpha\alpha}$, $C_{\alpha\beta}$*			
	sequence separation dependent	no	yes	yes	no			
	∡ between residue pairs	$\measuredangle C_\beta C_\alpha C_\beta$	no	no	no			
	width of distance intervals (Å)	1	1	0.25	1			
	maximum distance (Å)	≤14	∞	15	10			
	residue triplet interactions					○	○	
	hydrophobicity	○	○	○	○	○	○	○
	based on *accessible surface area*						✓	
	accessibility		✓			✓		✓
	number of neighbors	✓		✓	✓		✓	
	number of intervals	8	5	no limit	6	2	3	3
	polarity						○	
	hydrogen – bonds	○						
	local structure	○					○	
	sequence profile						○	
Probability normalization	on randomized sequences				✓	✓		
	on existing sequences	✓	✓	✓			✓	✓
	on the whole dataset	✓	✓	✓		✓	✓	✓
	per protein				✓			
Correction for sparse data		○	○	○				
# proteins in learning set		264	248	248	161	56	119	5

*The first 2 columns describe various terms of the potentials. The remaining columns indicate which of these terms were used by the predictor teams, referred to by their code (Table II). All the potentials summarized in this table have been derived from statistical analyses of known protein structures.

The **residue pair interactions** term is either a contact–noncontact potential, in which the definition of the interresidue contact is binary, or a distance-dependent potential, in which residue contacts are considered as a function of the spatial distance between the corresponding residues. The atoms used in computing the interresidue distances are indicated in the row *atoms considered.* $C_{\alpha\alpha}$ and $C_{\alpha\beta}$ indicate that a pseudoatom, positioned between consecutive C_αs, and between the C_α and C_β atoms of the same residue, were used, respectively. *Sequence separation dependent* refers to the practice of computing the residue pair interactions separately for different distances along the amino acid sequence. ∡ *between residue pairs* refers to the practice of subdividing interresidue distance intervals according to the angle between the vectors $C_\alpha(i)$–$C_\beta(i)$ and $C_\alpha(i)$–$C_\beta(j)$, where i and j are any two residues. The *width of distance intervals* and the *maximum distance* refer, respectively, to the interval and the largest distance (both in Å), considered in compiling residue pair probabilities from the dataset.

The **residue triplet interactions** term is computed from the propensities of 3 residues–contacts, in the protein database.

The **hydrophobicity** term is computed from the solvent exposure propensities of single residues. Such propensities are computed considering either, raw *accessible surface areas,* or *accessibilities.* Accessibility is evaluated either as the fraction of the total residue surface area which is exposed in the native protein, or from the *number of neighbors* of a given residue in the 3D structure. The number of intervals indicates the number of accessibility or of accessible surface area ranges used.

The **polarity** term is evaluated from the fraction of polar to nonpolar atoms surrounding a given residue. The **hydrogen bonds** term is evaluated from the propensities of residue pairs to be hydrogen bonded. The **local structure** term is computed from propensities of residues or residue pairs to adopt certain backbone dihedral angles, or a certain secondary structure. The **sequence profile** term is computed from the sequence similarity between target and threaded sequences.

Probability normalization refers to the schemes used to normalize database-derived probabilities. Normalization is performed either with respect to randomized sequences or to existing sequences.[1]

The **correction for sparse data** consists in giving a lesser weight to less frequent residues, residue pairs, or triplets, for which the statistics are less reliable.

The last row lists the number of proteins, protein chains, or structure motifs that were used to derive the potentials.

with the interactions at short distances representing the required contributions (Table V).

NIH seems to be the only group among those submitting alignments whose potential includes neither of these features. They also produce the only alignment with a random Q_3 score. An alternative explanation for this latter results could be that the alignment in question (ppdk3/1etu), corresponds to a random hit, and not to a statistically significant fold recognition.[27]

Interestingly, we find that the Q_3 scores are, on average, 10% lower and more often close to the random value for first hits corresponding to incorrect predictions (data not shown) than for the correct hits. The level of misprediction in these false positives is also significantly higher (~43% on average, compared to 16% on average for correct first hits). This indicates that the sequence–structure alignments in the false positives are often much more random than in the correct hits, suggesting in turn that, at least part of the time, threading is picking up a real signal.

Does threading tend to maximize hydrophobic interactions in the protein core?

The last column of Table IV compares the percentage of hydrophobic residues buried in the core of the hit structure when the target sequence is aligned to it to the same quantity computed with the native hit sequence. We see that, on average, the alignments predicted by threading bury more or less the same fraction of hydrophobic residues in the protein core as the native hit protein, or for that matter as the native target protein (data not shown).

It appears, however, that some of the poorly predicted spatial alignments bury a significantly larger fraction of hydrophobic residues in the core than the corresponding native hit protein. For instance, the kau B/2hre alignment from London, in which none of the secondary structure elements is correctly positioned in space, yields a protein core in which 75% of the buried residues are hydrophobic. In the native 2rhe protein only 64% of the buried core residues are hydrophobic, which is 11% less. The same trend is observed in other poor alignments, such as those of synapto/1cd8 from London, prosub/2fxb from Salzburg, and kau B/2fjb L from Oxford. These alignments yield protein cores in which 80, 87, and 75% of the residues are hydrophobic, respectively. In comparison, the fraction of hydrophobic residues in the corresponding native cores is only 69, 67, and 57%, respectively.

These results suggest that the analyzed threading procedures tend to maximize the hydrophobic interactions in the protein core. This could have the effect of reducing the number of possible conformations compatible with the target sequence, thereby leading to fold recognition, even in the absence of correct spatial alignment. This tendency might be strong enough to produce scrambled sequence–structure alignments with a nonnative character. A brief analysis of the false positives shows that there too, the fraction of hydrophobic residues buried in the protein core remains rather high and similar to the fractions observed in native proteins. This underscores the important, and possibly too important, role played by hydrophobic interactions in many of the threading potentials (see below), and may explain why the scores for correct and incorrect hits are, in general, indistinguishable.

It is noteworthy that the poor spatial alignment obtained by Cambridge for the target/hit pair of staufen3/2cmd, which also corresponds to a correct recognition, displays opposite behavior. It buries a much lower fraction of hydrophobic residues (37%) in the core than the native 2cmd protein (63%). In this case, fold recognition is clearly not fostered by optimizing hydrophobic interactions, in agreement with the fact that the Cambridge approach relies essentially on scoring against sequence profiles (see below).

Are Some Methods More Effective Than Others?

Summary of the prediction methods

An attempt at summarizing the different aspects of the threading approaches is made in Tables V and VI. In all these approaches, the sequence of the target protein is mounted onto an ensemble of known structures, and effective potentials derived from known protein structures are used to measure the quality of the sequence–structure matches.

Most of the potentials (Table V) contain contributions derived from database frequencies of residue–residue contacts or of distance-dependent interactions between residue pairs. Hydrophobic interactions play a major role in these contributions.[28,29] Some predictors add terms based on residue solvent exposure propensities, which also takes into account the hydrophobic effect, as well as terms which consider local backbone conformational preferences or on hydrogen bonds.

A majority of the threading methods (Table VI) involve classical dynamic programming algorithms. Those are usually applied without updating the residue environment to take into account the presence of the mounted sequence (the frozen approximation). The exceptions are the London group, which applied a double dynamic programming method, and the NIH group, which used a Monte Carlo procedure. In most cases insertion and deletion penalties are applied. In general, threading is performed on full motifs, but in some cases (e.g., the NIH group) only secondary structures in the core are considered.

Different approaches were used by Oxford and Cambridge. The approach of Oxford may be regarded as midway between ab initio and threading predictions.[30] In this approach, sequences homologous to the target are collected and aligned by a combination of automatic procedures and hand editing. From the multiple alignment, a consensus secondary structure prediction is derived. This consensus prediction is then matched to the secondary structures of proteins in the database, using various scoring schemes. Highly scoring matches are filtered based on compatibility with observed β-strand contacts and on their ability to satisfy constraints on loop lengths. Cambridge uses multiple alignments of sequences homologous to the target, to build a Hidden Markov Model. This model is then used to search for highly scoring sequences from among a representative set corresponding to known protein structures. These sequences are then reranked and

TABLE VI. Summary of the Threading Methods Used by the Predictors*

		Osaka	London	Salzburg	NIH	Scripps	EMBL	Baltimore
Threading method	**dynamic programming**	○	○	○		○	○	○
	simple, with native environment	✓		✓				✓
	simple, with environment update					✓	✓	
	double		✓					
	Monte Carlo				○			
Indel penalty	**indel penalty**	○	○	○		○	○	○
	indels in secondary structures		✓	✓		✓	✓	✓
	indels of secondary structures	✓	✓	✓		✓	✓	✓
	lower/upper bound on loop length				○			
Threading over	whole structure	✓	✓	✓		✓	✓	✓
	core motif		✓		✓			
# proteins or cores in motif library		325	242–265	288/1265	435	381	?	105

*Columns 1 and 2 describe the different methods. The other columns indicate which predictor teams used them. The predictors are referred to by their codes (Table II). The threading methods use either **dynamic programming** (the most common approach), or **Monte Carlo** procedures (used only by the NIH group). Different variants of the dynamic programming method are used for threading. The simplest uses straightforward dynamic programming algorithms, taking into account only the 3D environment in the native protein. Another performs a number of iterations in which the protein 3D environment is updated, to consider the presence of the threaded sequence. Double dynamic programming combines simple dynamic programming runs, performed for each threaded residue fixed on a given position in the structure.[38] **Indel penalty** refers to the practice of penalizing insertions and deletions. Some authors impose instead **lower and upper bounds on loop lengths.** Insertions and deletions may or may not be allowed inside secondary structures elements. Also insertions and deletions of complete secondary structure elements may or may not be allowed. Threading may be performed on the full structure, or on portions of the structure, for example, the protein core. London combines both procedures; their residue pair interaction is evaluated only for residues in secondary structures, but the hydrophobicity term is computed for all the residues. The number of known protein structures or core motifs onto which each team performed the threading operation is given in the last row. London used sets varying in size between 242 and 265. Salzburg used 288 proteins for some predictions and 1265 for others. '?' indicates that the information was unavailable.

filtered,[31] using information on secondary structure and β-strand pairing predictions, and on fold classification.

Linking aspects of the prediction methods to their performance

It would be most useful to be able to rationalize the results of this contest in terms of specific aspects of the prediction methods summarized above. This would require, however, a detailed knowledge of the different computational approaches, as well as additional data on the computational results, which were unfortunately not available to us. It would also require a systematic investigation in close collaboration with the predictors in which key parameters of the methods are varied one at a time. In what follows, therefore, we highlight only some of our impressions, which should be considered as preliminary at best.

Aspects of the effective potentials. It seems to us that better prediction performance can be linked to more specific potentials, which combine moreover several contributions, from local and nonlocal interactions between residues along the sequence, as well as from solvation effects. The latter aspect is in agreement with previous results on native fold recognition.[32] This is the case for the Salzburg, London, and to some extent Osaka groups. The potentials of the former 2 groups feature a very fine sampling of the residue separation along the sequence. Salzburg uses, moreover, fine sampling of the interresidue spatial distance. The residue pair potential of Osaka has no dependence on sequence separation, but it includes an angular dependence. It is furthermore supplemented by terms representing backbone dihedral angles preferences and H-bond propensities. The very crude residue contact potential of Scripps, although combined with residue triplets interactions, seems to fare less well. As noted above, neglecting altogether contributions from local interactions, as the NIH potential seemingly does, seems to be a disadvantage, although it is admittedly not possibly to generalize from a single example.

Aspects of the threading algorithms. The performance of the threading algorithms is even harder to evaluate under the present circumstances. Visual inspection of the alignments suggests that the gap penalties may not always be appropriate. We saw, indeed, that in some cases the gaps were obviously far too small. Interestingly, NIH was the only group that did not introduce gap penalties. They only imposed lower and upper limits on loop lengths. This group also threads exclusively onto secondary structure elements composing the protein core and not onto the full structure, as virtually all the other groups do. These features are interesting, even though their current implementation does not seem to be accompanied by a better performance.

It is rather puzzling that the simple dynamic programming algorithms without environment update do not yield lower performance scores than the more sophisticated algorithms, which update the 3D environment so as to take into account the presence of the mounted sequence.

Other aspects. The approaches by the Oxford and Cambridge groups, which rely on sequence information from homologous proteins and use consensus secondary structure predictions, seem to perform rather well. This impression rests, however, on a very small sample of 2 and 4 predictions from each group, respectively, and should therefore not be generalized too quickly. It is noteworthy that both groups also used various additional filters. Oxford, in particular, employed manual intervention, in which expert knowledge plays an important role. It is thus the sum of this knowledge and the automatic procedures which is being assessed in this case.

However, manual editing of the fold predictions obtained by automatic procedures may have pitfalls, as seen from the following example. The threading procedure of Baltimore correctly identified the viral coat protein (2btv) as the top ranking fold for the target synapto. But visual inspection led the predictors to choose the second hit, hemagglutinin (1hmg), as the correct fold, thereby making a wrong prediction.

CONCLUDING REMARKS

The prediction results reviewed here show that current threading methods are capable of identifying the correct fold in many cases, even in absence of detectable sequence homology between the target and any protein of known structure. This performance is particularly encouraging since it was obtained even in cases where the degree of structural similarity between the target and the known folds is rather low.

Our assessment shows, however, that the fold recognition problem is not solved, and that there is quite some room for improvement. In particular, the sequence–structure alignments are, in general, quite poor, and means to distinguish between the correct fold and false positives still seem to be largely lacking. Since many of the assessed prediction methods are still in the development stage, we should expect some of these improvements in the very near future.

The observation that every method predicts correctly a different set of targets suggests that it could be worthwhile to combine aspects of different methods. One could also imagine using consensus predictions which combine the results of several available methods, in much the same way as it is being done for secondary structure predictions.[33,34] Such consensus predictions would be time consuming, but much more reliable than any of the methods taken individually. Threading methods could also profit from the use of secondary structure predictions based on sequence alignments,[35] since these have greatly boosted the accuracy of secondary structure predictions in recent years.[36]

It is necessary, however, to point out, once more, important limitations of our evaluation. It did not consider about half of the submitted predictions, namely those made for targets which turned out not to correspond to known folds. Furthermore, false positives were analyzed only very briefly, and lower ranking predictions were not examined at all.

Lastly, we would like to mention that one additional team submitted predictions for 2 targets, which were based entirely on conventional sequence alignment techniques. None of these predictions identified the correct fold. Furthermore, applying FASTA[37] to align the sequences of the 11 targets with known folds to the sequences of our representative protein set, we were able to verify that it yielded essentially random hits. The highest scoring hits had, moreover, no obvious relationships with one another. Hence, when there is no significant sequence homology between the target and the known folds, threading methods appear to be clearly superior to conventional sequence alignments. This is in itself a quite significant result, which should encourage us to pursue the development of this exciting approach.

ACKNOWLEDGMENTS

All the predictors deserve praise for taking part in this contest, in particular, for subjecting their approaches to public evaluation and for the massive work they accomplished under severe time constraints. We furthermore thank the predictors for their cooperation and help, the crystallographers for making available their structures prior to publication, and John Moult, for putting everyone to a challenging test. Without them, this work would not have been possible. We acknowledge Nathalie Boutonnet and Jean Richelle for valuable help with the 3D structure alignments, and with the computer systems, respectively. M.J.R. is Research Associate at the National Fund for Scientific Research (Belgium), and C.L. is Research Assistant at the Fonds pour la Formation à la Recherche dans l'Industrie et dans l'Agriculture (FRIA). The Belgian programme of Interuniversity Poles of Attraction initiated by the Belgian State, Prime Minister's Office for Science and Technology and Culture, is gratefully acknowledged for support.

REFERENCES

1. Bryant, S.H., Altschul, S.F. Statistics of sequence–structure threading. Curr. Opin. Struct. Biol. 5:237–244, 1994.
2. Wodak, S.J., Rooman, M.J. Generating and testing protein folds. Curr. Opin. Struct. Biol. 3:247–259, 1993.
3. Fetrow, J.S., Bryant, S.H. New programs for protein ter-

tiary structure prediction. Biotechnology 11:479–484, 1993.

4. Sippl, M.J. Knowledge-based potentials for proteins. Curr. Opin. Struct. Biol. 5:229–235, 1995.
5. Jones, D., Thornton, J.M. Protein fold recognition. Comp. Aided Mol. Design 7:439–456, 1993.
6. Cothia, C., Lesk, A.M. The relation between the divergence of sequence and structure in proteins. EMBO J. 5:823–826, 1986.
7. Chothia, C., Lesk, A.M. The evolution of protein structures. Cold Spring Harbor Symp. Quant. Biol. LII:399–405, 1987.
8. *Hubbard*, T.J.P., Blundell, T.L. Comparison of solvent-inaccessible cores of homologous proteins: Definitions useful for protein modelling. Prot. Eng. 1:159–171, 1987.
9. Holm, L., Ouzounis, C., Sander, C., Tuparev, G., Vriend, G. A database of protein structure families with common folding motifs. Prot. Sci. 1:1691–1698, 1992.
10. Orengo, C.A., Flores, T.P., Taylor, W.R., Thornton, J.M. Identification and classification of protein fold families. Prot. Eng. 6:485–500, 1993.
11. Johnson, M.S., Overington, J.P., Blundell, T.L. Alignment and searching for common protein folds using a data bank of structural templates. J. Mol. Biol. 231:735–752, 1993.
12. Chothia, C. One thousand protein families for the molecular biologist. Nature (London) 357:543–544, 1992.
13. Orengo, C.A., Jones, D.T., Thornton, J.M. Protein superfamilies and domain superfolds. Nature (London) 372:631–634, 1994.
14. Jones, D.T., Taylor, W.R., Thornton, J.M. A new approach to protein fold recognition. Nature (London) 358:86–89, 1992.
15. Godzik, A., Kolinski, A. Sequence–structure matching in globular proteins: Application to supersecondary and tertiary structure predictions. Proc. Natl. Acad. Sci. U.S.A. 89:12098–12102, 1992.
16. Bernstein, F., Koetzle, T., Williams, G., Meyer, E., Brice, M., Rodgers, J., Kennard, O., Shimanoushi, T., Tasumi, M. The Protein Data Bank: a computer-based archival file for macromolecular structures. J. Mol. Biol. 112:535–542, 1977.
17. Boutonnet, N.S., Rooman, M.J., Ochagavia, M.-E., Richelle, J., Wodak, S.J. Optimal protein structure alignments by multiple linkage clustering: application to distantly related proteins. Prot. Eng. 8, 1995 (in press).
18. Boutonnet, N.S., Rooman, M.J., Wodak, S.J. Automatic analysis of protein conformational changes by multiple linkage clustering. J. Mol. Biol., 1995 (in press).
19. Kabsch, W., Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22:2577–2637, 1983.
20. Murzin, A.G., Brenner, S.E., Hubbard, T., Chothia, C. SCOP: A structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol. 247:536–540, 1995.
21. Rost, B., Schneider, R., Sander, C. Redefining the goals of protein secondary structure prediction. J. Mol. Biol. 235: 13–26, 1994.
22. Alard, P. Computations of surface areas and energies in the field of macromolecules. Ph.D. Thesis, Université Libre de Bruxelles, Belgium, 1990.
23. Orengo, C.A. Classification of protein folds. Curr. Opin. Struct. Biol. 4:429–440, 1994.
24. Kong, X.P., Onrust, R., O'Donnell, M., Kuriyan, J. Three-dimensional structure of the beta subunit of E. coli DNA polymerase III holoenzyme: A sliding DNA clamp. Cell 69:425–437, 1992.
25. Kurijan, J., O'Donel, M. Sliding Clamps of DNA Polymerases. J. Mol. Biol. 234:915–925, 1993.
26. Presnell, S.R., Cohen, B.I., Cohen, F.E. A segment-based approach to protein secondary structure prediction. Biochemistry 31:983–993, 1992.
27. Madej, T., Gibrat, J.-F., Bryant, S.H. Threading a database of protein cores. 1995 (submitted).
28. Casari, G., Sippl, M.J. Structure-derived hydrophobic potential. Hydrophobic potential derived from X-ray structures of globular proteins is able to identify native folds. J. Mol. Biol. 224:725–732, 1992.
29. Bryant, S.H., Lawrence, C.E. An empirical energy function for threading protein sequence through folding motif. Proteins Struct. Funct. Genet. 16:92–112, 1993.
30. Russell, R.B., Copley, R.R., Barton, G.J., Protein fold recognition from secondary structure assignments. In: "Proceedings of the 28th Annual Hawaii International Conference on System Sciences," Vol. 5. New York: IEEE Press, 1995:302–311.
31. Hubbard, T.J., Park, J. Fold recognition and ab initio structure predictions using Hidden Markov Model and β-strands pair potentials. Proteins, 1995 (in press).
32. Kocher, J.-P.A., Rooman, M.J., Wodak, S.J. Factors influencing the ability of knowledge based potentials to identify native sequence-structure matches. J. Mol. Biol. 235: 1598–1613, 1994.
33. Biou, V., Gibrat, J.F., Levin, J.M., Robson, B., Garnier, J. Secondary structure prediction: Combination of three different methods. Prot. Eng. 2:185–191, 1988.
34. Zhang, X., Mesirov, J.P., Waltz, D.L. Hybrid system for protein secondary structure prediction. J. Mol. Biol. 225: 1049–1063, 1992.
35. Russel, R.B., Sternberg, M.J.E. Structure prediction—how good are we? Curr. Biol. 5:488–490, 1995.
36. Rost, B., Sander, C. Prediction of protein secondary structure at better than 70% accuracy. J. Mol. Biol. 232:584–599, 1993.
37. Sellers, P. On the theory and computation of evolutionary distances. SIAM J. Appl. Math. 26:787–793, 1974.
38. Taylor, W.R., Orengo, C.A. Protein structure alignment. J. Mol. Biol. 208:1–22, 1989.
39. Kraulis, P.J. MOLSCRIPT—a program to produce both detailed and schematic plots of protein structures. J. Appl. Cryst. 24:946–950, 1991.
40. Bacon, D.J., Anderson, W.F. A fast algorithm for rendering space-filling molecule pictures. J. Mol. Graph. 6:219–220, 1988.
41. Meritt, E.A., Murphy, M.E.P. Raster3D version 2.0: A program for photorealistic molecular graphics, 1994 (submitted).

APPENDIX: ATLAS OF TARGET PROTEINS FOR THE ASILOMAR THREADING CONTEST, AND THE KNOWN FOLD THEY MOST RESEMBLE

Figure A1 depicts the target structures for which threading predictions submitted to the Asilomar contest were assessed, alongside the structures they most resemble. These latter structures were identified by an automatic structure alignment procedure from a data set of 203 representative folds (see The Representative Set of Known Protein Folds and Table Ia). The 3D alignment procedure was the MLC version of SoFiSt (see Structure Alignment Methods). Each individual part of the figure (denoted a through k) refers to a particular target-known fold pair. The segments of secondary structures (blocks) aligned by SoFiSt are shown in yellow, α-helices are depicted in green, β-strand in blue, and coils in purple. Secondary structure assignments were computed using DSSP.[19] When only a portion of a β-strand was aligned by SoFiSt, the corresponding portion is depicted as an independent β-strand. Pictures were obtained using the programs MOLSCRIPT[39] and Raster-3D.[40,41] Permission was denied to reproduce pictures of the targets ppdk3, ppdk4, and pbdg, as the corresponding X-ray structures have not yet been published.

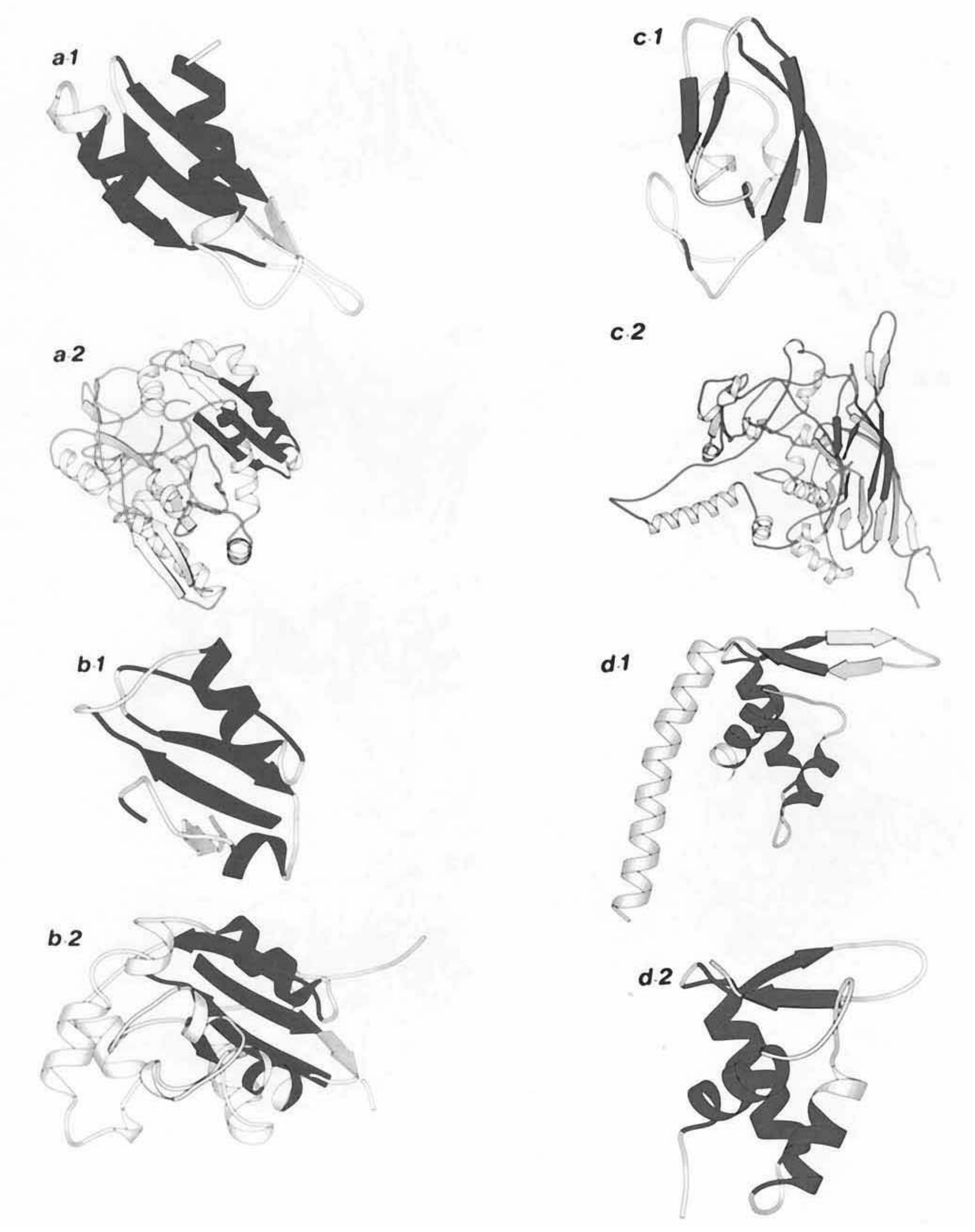


Fig. A1 (a–d).

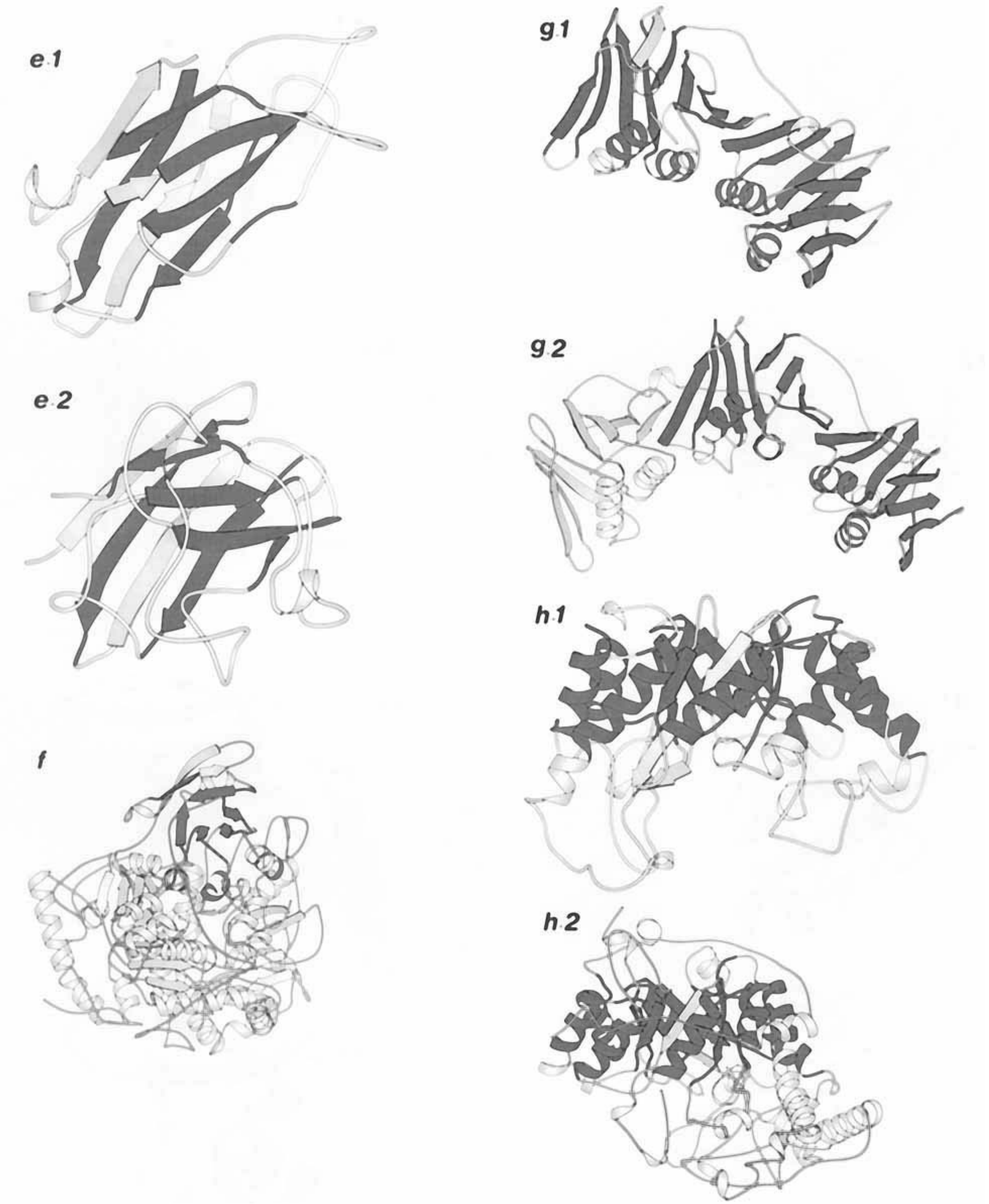
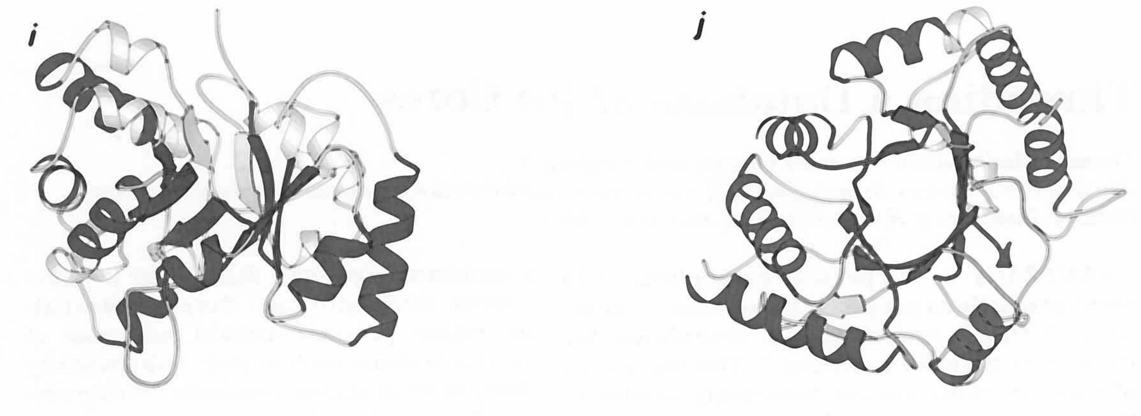


Fig. A1 (e–h).



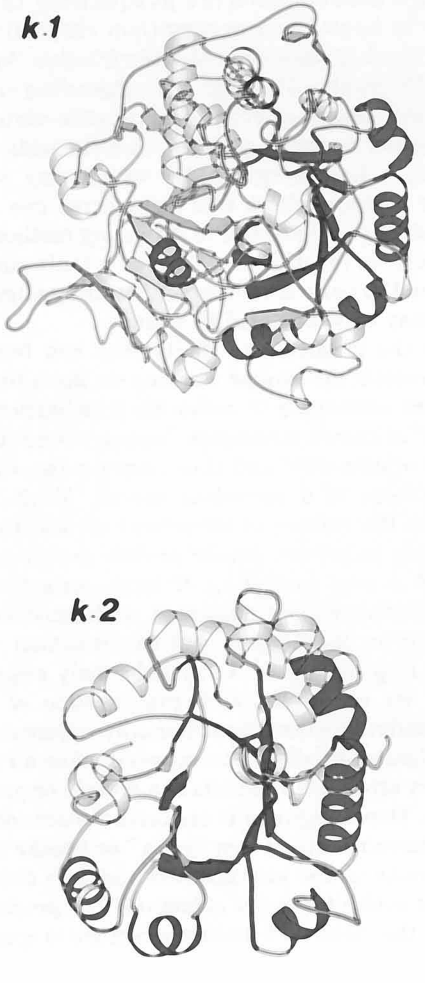


Fig. A1 (i–k).

Fig. A1. (a1) Target: staufen3 (domain 3 of staufen from *Drosophila*). NMR structure determined by M. Bycroft. (*a2*) Known fold: 1pda (porphobilinogen deaminase). (b1) Target: prosub (propeptide from subtilisin BPN'). X-Ray structure determined by T. Gallagher et al. (b2) Known fold: 2nck L (nucleoside diphosphate kinase). (c1) Target: urease chain B, from *Klebsiella aerogenes.* X-Ray structure determined by E. Jabri and A. Karplus. (c2) Known fold: 2bpa L (bacteriophage ϕX174 capsid protein). (d1) Target: rtp replication terminator protein from *Bacillus subtilis.* X-Ray structure determined by D. Bussiere and S. White. (d2) Known fold: 1hst A (histone H5 globular domain). (e1) target: synapto (synaptotagmin, first C2 domain). X-Ray structure determined by S. Sprang. (e2) Known fold: 1cob A (superoxide dismutase). The depicted SoFiSt structure alignment of synapto and 1cob A unduely misses detecting the similarity of the 2 β-strands, spanning residues 157–166 of synapto and A26–A34 in 1cob A, due to the presence of a β-bulge at position 163 of synapto, as the latter increases the pairwise segment rms value to 1.8 Å. (f) 1aco (aconitase), the known fold identified by SoFiSt as most resembling the target ppdk3 (pyruvate phosphate dikinase, domain 3, residues 370–520), whose X-ray structure has been determined by O. Herzberg. (g1) Target: DNA polymerase processivity factor from yeast. X-Ray structure determined by T. Krishna et al. (g2) Known fold: 2pol A (polymerase III from *E. coli*). (h1) Target: xylanase (catalytic core) from *Pseudomonas fluorescences.* X-Ray structure determined by G.W. Harris et al. (h2) Known fold: 1btc (β-amylase). (i) 5tim (triose phosphate isomerase), the known fold identified by SoFiSt as most resembling the target ppdk4 (pyruvate phosphate dikinase, domain 4, residues 520–874), whose X-ray structure has been determined by O. Herzberg. (j) 1nar (narbonin), the known fold identified by SoFiSt as most resembling the target pbdg (6-phospho-β-D-galactosidase from *Lactococcus lactis*), whose X-ray structure has been determined by C. Wiesmann. (k1) Target: kau A (urease chain A from *Klebsiella aerogenes.* X-Ray structure determined by E. Jabri and A. Karplus. (k2) Known fold: 5tim (triose phosphate isomerase).