

EVALUATING PREDICTIONS OF SECONDARY STRUCTURE IN PROTEINS

Thomas F. Jenny and Steven A. Benner

Department of Chemistry, E.T.H., CH-8092 Zurich, Switzerland

Received February 16, 1994

To learn how secondary structure assignments diverge during divergent evolution, pairs of proteins with solved crystal structures were aligned and their assignments compared as a function of evolutionary distance. Residues assigned in one structure to a helix or a strand are frequently paired with residues assigned in the other to a coil. However, residues assigned to a helix in one structure are almost never paired with residues assigned to a strand in the other. This suggests additional limitations to the "three state residue-by-residue" score commonly used to evaluate secondary structure predictions and suggests recommendations for how secondary structure predictions should be scored to assess accurately their value as starting points for modelling tertiary structure.

© 1994 Academic Press, Inc.

In the past five years, new tools have emerged that predict the secondary structure of proteins by analyzing the patterns of conservation and variation within a set of aligned homologous sequences (1, 2). These were used independently to make *bona fide* predictions (those announced before an experimental structure becomes available) for the secondary structure of the cytokine receptor and protein kinase protein families (3, 4). Subsequently determined crystal structures of members of the receptor and kinase families respectively showed that both predictions were remarkably accurate (5, 6). As with all methods that predict secondary structure from a set of aligned homologous sequences, these predictions are *consensus models*, applying to a family of proteins as a whole, not to any individual member of the family. The approach has since been used to predict consensus models for the SH2 domain (7, 8), the SH3 domain (9, 10), the hemorrhagic metalloproteases (11), and other protein families (1,12):

This approach complements other approaches that also start from a set of aligned protein sequences but do not explicitly analyze patterns of conservation and variation within that set. One particularly popular approach averages secondary structure predictions made by classical methods (13, 14) for individual members of a protein family. This approach was suggested in the 1970's (14, 15, 16), but first applied in a *bona fide* prediction setting only in the mid 1980's (17, 18). The approach has been developed by Taylor (19), Sternberg (20), Barton (21) and their coworkers, where predictions have often been aided by circular dichroism data, pattern identification, or other independent indicators of secondary structure, with some impressive results. These have complemented some significant *de novo* predictions (22), and some new neural networks (23, 24) that improve upon their predecessors (25) by using multiple aligned sequences as input.

With the availability of high quality predictions of secondary structure starting from sequence data alone, two new issues have emerged. The first asks how the elements of a predicted secondary structure might be assembled to yield a model for tertiary structure. The second asks how high-quality secondary structure predictions should be judged to assess its value as the starting point for modelling tertiary structure. These issues are related, as evaluations of secondary structure predictions made by various methods will determine in part which methods are used to predict secondary structures as the starting point for modelling tertiary structure.

We have compared the divergence in the experimentally assigned secondary structures of homologous protein pairs as a function of evolutionary distance. This comparison confirms that secondary structure assignments for homologous protein with quite similar overall tertiary structures can diverge significantly (2). This shows not only that special tools are needed for evaluating consensus predictions for secondary structure (2), but also that a secondary structure prediction can contain some types of errors (but not other types) without compromising its value as the starting point for modelling tertiary structure. This in turn suggests that the tools commonly used to evaluate secondary structure predictions do not measure their true value for modelling tertiary structure, and suggests new approaches for scoring secondary structure predictions that do.

METHODS

A series of pairwise alignments of protein sequences having secondary structures assigned to crystallographic coordinates using the DSSP (Define Secondary Structure of Proteins) (26) method were extracted from SwissProt Version 26 (27) using the DARWIN system. Such alignments are used routinely in Zürich for structure prediction work. Both the pairwise alignments and the computer tools used to generate them are available by electronic mail by server (at the address cbrg@inf.ethz.ch). The aligned pairs were then ranked by evolutionary distance measured in PAM units (28), the number of Point Accepted Mutations per 100 amino acids (80 PAM = 48% pairwise identity; 100 PAM = 41% pairwise identity; 120 PAM = 36% pairwise identity).

Two sets of proteins were examined, the first with crystal structures with resolutions better than 2Å, the second with resolutions of better than 3Å. Concatenated proteins and crystal structures of proteins associated with membrane-bound units were excluded from this analysis. Based on the sequence alignment, secondary structural assignments for the two homologous proteins were aligned and compared, and then scored residue-by-residue by type (alpha helix paired with helix, strand, coil, or gap, beta strand paired with strand, coil or gap, and coil paired with gap).

RESULTS

Table 1 records the divergence in secondary structural assignments made by DSSP (26) as a function of evolutionary distance. Several features of the results are worth noting.

First, residues participating in a helix are essentially never paired in a homologous protein with residues participating in a strand. The only significant exception occurs at very high PAM distances.

Second, pairing of residues assigned to a helix or strand in one structure with residues assigned to a coil in the other occurs to a significant extent, more frequently in structures with low resolution, but (surprisingly) not significantly more frequently in proteins with higher evolutionary divergence.

Further, the pairing of a residue assigned to a helix, strand or coil with a gap increases with increasing evolutionary divergence, as expected based on empirical studies of the appearance of

Table 1. Pairing of Secondary Structural Assignments in Homologous Proteins

PAM Window	Total sequence pairs	Total positions	h-h % norm†	s-s % norm†	c-c % norm†	h*s %**	h*c %	s*c %	h*g %	s*g %	c*g %			
(a) Protein pairs with High Resolution (< 2 Å) Crystal Structures (145 proteins)														
000-020	7	1319	37	96	10	93	49	96	0	3	1	0	0	0
020-040	4	573	22	88	14	94	55	93	0	6	2	0	0	1
040-060	9	744	15	95	16	92	63	95	0	2	2	0	0	2
060-080	9	1618	14	87	28	92	46	90	0	4	4	0	0	3
080-100	11	1773	28	82	12	84	40	82	0	11	3	1	1	3
100-120	11	2149	14	78	27	91	42	86	0	6	4	2	1	4
120-140	20	2920	18	85	23	86	38	81	0	6	6	1	1	6
140-160	16	2791	22	82	21	83	34	77	0	8	7	1	1	5
160-180	13	2076	23	84	22	84	33	78	0	7	7	1	1	4
180-200	13	1986	45	83	6	79	22	66	0	16	3	3	1	4
200-220	0	0	0	0	0	0	0	0	0	0	0	0	0	0
(b) Protein pairs with Medium Resolution (< 3 Å) Crystal Structures (255 proteins)														
000-020	24	5224	39	89	7	89	41	88	0	10	2	0	0	0
020-040	15	2540	23	89	18	88	48	90	0	6	4	0	0	1
040-060	21	3859	20	83	17	88	47	87	0	7	4	1	0	3
060-080	28	4615	30	90	18	88	39	86	0	6	5	1	0	2
080-100	32	5052	35	84	9	78	35	78	0	11	4	2	1	4
100-120	36	5984	21	85	21	85	39	82	0	6	6	1	1	4
120-140	35	5275	18	85	23	85	39	82	0	5	6	1	1	6
140-160	36	5382	28	85	16	81	33	78	0	8	6	2	1	5
160-180	38	5333	46	86	9	83	24	72	0	13	3	1	0	3
180-200	34	5303	51	85	4	70	22	68	0	15	3	2	1	2
200-220	11	1596	41	82	5	57	28	72	2	14	5	2	0	2

* 12 positions in the entire data set match a helix residue with a strand residue.

** 84 positions in the entire data set match a helix residue with a strand residue.

% records as a percentage the number of positions in the data set having the indicated pairing divided by the total number of positions paired.

† normalized scores report, as a percentage, twice the number of positions that pair a particular secondary structure type in one sequence with the same type in the other (x*x), divided by the total number of positions where that secondary type is found in one of the two proteins.

PAM window: The evolutionary distance separating the two protein sequences paired, measured in the number of Point Accepted Mutations per 100 amino acids (28).

h-h: Alpha helix assignment paired with an alpha helix assignment.

s-s: Beta strand assignment paired with a beta strand assignment.

c-c: Coil assignment paired with a coil assignment.

h*s: Alpha helix assignment paired with a beta strand assignment.

h*c: Alpha helix assignment paired with a coil assignment.

s*c: Beta strand assignment paired with a coil assignment.

h*g: Alpha helix assignment paired with a gap.

s*g: Beta strand assignment paired with a gap.

c*g: Coil assignment paired with a gap.

gaps in protein alignments during divergent evolution (29, 30). Further, coils are most frequently paired with gaps, an observation now exploited by most structure prediction methods.

Finally, in each of the homologous protein pairs, the overall tertiary structure is essentially the same, certainly up to PAM distances of ca. 150 (31, 32). Yet the secondary structural assignments differ. For example, in the PAM 100-120 window, 10-12% of the helix and strand secondary

structural assignments made for one protein do not correspond to helix and strand assignments in the homolog. This cannot be due to different assignment heuristics, as all assignments were made using the DSSP method (26). Nor do these arise through misalignment; only 3.8% percent of the positions are expected to be misaligned by DARWIN at this PAM distance (33), and the number of secondary structural mismatches arising from these misalignments should be much less than this. Thus, at a PAM distance of 100-120 (corresponding to two protein sequences that are ca. 40% identical), two proteins with few if any meaningful differences in global structure will generally have secondary structure assignments that differ at ca. 10% of the positions.

DISCUSSION

One tool often used to evaluate predicted secondary structures is the "three state residue-by-residue score" (34, 35). To calculate this score, each residue in a protein is assigned to one of three states: alpha helix, beta strand, and coil (neither helix nor strand). The experimental data are then used to make analogous experimental assignments to each residue. The three state score is calculated by dividing the number of correct assignments [(helix paired with helix) + (strand paired with strand) + (coil paired with coil)] by the total number of assignments; the result is reported as a percentage. The scores for many predictions are often combined to yield an average score, which is frequently believed to be a useful figure of merit for a prediction method (23, 24; 35).

Despite its widespread use, the three state residue-by-residue score is problematic. First and most simply, there is no universally accepted way to assign an experimental secondary structure to a set of crystallographic data. Thus, the experimental standard used to evaluate a prediction is itself somewhat arbitrary. Colloc'h *et al.* (36) recently pointed out that the assignments of secondary structure made by three computer tools agreed at only 63% of the positions. This implies that a per residue score cannot routinely be higher than ca. 75% ($63\% + 37\%/3$), as a better score obtained by comparison with an experimental assignment made using one tool will be worse when the comparison is made with an experimental assignment made using another tool.

Even if a fully objective tool were available for assigning experimental secondary structures, however, the three state score would remain problematic as an indicator of the value of a prediction as the starting point for building a tertiary structural model. This is so because the score treats different types of errors identically. For example, the misassignment of a residue as part of an alpha helix when it in fact forms part of a beta strand has the same impact on the three state score as the misassignment of a residue as part of a coil when it is in fact part of a beta strand. Yet, as Table 1 shows, the second type of misassignment has essentially no impact on overall tertiary structure; it occurs frequently in homologous proteins that have essentially identical tertiary structures. In contrast, misassignment of a strand as a helix (or vice versa) has a significant impact on a tertiary structure model, and essentially never occurs when comparing two homologous proteins. Yet the three state score counts the two types of errors the same.

This means that two predictions may both have 90% three state residue-by-residue scores, but quite different values as starting points for building a tertiary structural model. If the 10% incorrect assignments confuse helix and strand residues for coil residues (or vice versa), the model could be perfectly adequate for modelling tertiary structure, especially if the mistakes are made at the ends of

helices and strands. If, however, the 10% incorrect assignments confuse helix residues for strand residues and vice versa, it almost certainly could not.

This was clearly illustrated in the recent prediction contest for the hemorrhagic metalloprotease family (11), where consensus predictions were made from seven homologous protein sequences. Using subsequently determined crystal structures (37, 38), the three state per residue scores of the predictions produced by the ETH method (1, 2) and the neural network developed by Sander and coworkers in Heidelberg (24) were 71.3 and 66.3% respectively, not greatly different. However, one third of the errors made in the Heidelberg prediction confused residues in alpha helices with residues in beta strands; 11% of the sequence was therefore seriously misassigned. In contrast, the ETH prediction seriously misassigned only 2 residues (<1%). Thus, the ETH prediction was better than the Heidelberg prediction as the starting point for modelling the tertiary structure of the protein, even though the three state residue-by-residue scores differed by only 5 percentage points. A similar feature was noted in the ETH prediction for protein kinase (4, 39).

The three state score becomes still more problematic when evaluating a consensus prediction for a family of proteins. When predicting a secondary structure from an alignment of protein sequences, one assumes that homologous proteins have exactly the same secondary structure. Yet, as Table 1 shows, this is true only as an approximation. The Table suggests that for proteins ca. 120 PAM units divergent, a three-state per-residue score for a consensus prediction cannot be much higher than 90%. A higher score arising from a comparison of the predicted secondary structure with an experimental structure for one family member implies a lower score when the comparison is made with a structure for another, making the score depend arbitrarily on which protein in the family happens to yield the first crystal structure.

Recently, Russell and Barton (40) proposed to revise the three state scoring approach in light of the last difficulty. Their proposal, to diminish the target score for a "perfect" prediction when evaluating a consensus prediction and to identify "core" secondary structure when more than one experimental structure is available, is reasonable and should be adopted. However, their revision does not address the first two problems: the non-objectivity of experimental secondary structure assignments, and the different impact of different types of errors on the value of a secondary structure prediction as a starting point for building a tertiary structural model. Thus, the approach does not repair the most serious flaws in the three state scoring method.

Instead, both our results and the results reported by Russell and Barton (40) suggests that no single number adequately describes the value of a secondary structural prediction as the starting point for modelling tertiary structure. This implies that efforts to obtain such abbreviated scores are not likely to contribute to the development of methods for predicting the conformation of proteins. Rather, scoring methods should reflect the fact that strategies for modelling tertiary structure from secondary structure predictions are not yet well developed. To this end, we make the following recommendations:

(a) Residue-by-residue scores that reflect the number of residues assigned correctly should always be accompanied by scores that reflect the number of serious errors, those where helix residues are mistaken for strand residues and vice versa. Such errors that will most seriously obstruct assembly of a tertiary structural model from a predicted secondary structure.

(b) To be most useful for the development of methods, correct assignments made for helices, strands, and coils should be reported separately, as in Table I (but where predicted and experimental structures rather than two experimental structures are compared). Different methods assign different types of secondary structural units with different accuracies. The best prediction package may ultimately use different methods for different parts of the prediction, and it is important to preserve this information from individual predictions.

(c) In any case, scores calculated residue-by-residue are most meaningful when accompanied by evaluations made by individual secondary structural elements (2, 41).

(d) When evaluating a consensus prediction, multiple experimental structures are best used, where a "core" experimental secondary structure is calculated from these and used fairly to evaluate the consensus prediction. When multiple experimental structures are not available, the target score for a "perfect" prediction should be lowered to reflect divergence in secondary structure.

(e) In any case, to compare different methods, the predictions must be made under similar conditions starting from similar data. Recent papers (40, 41) have attributed different merits to different prediction methods by comparing predictions made by one method exclusively from sequence data with predictions made by another where circular dichroism data or other experimental indicators of secondary structure contributed to predictions. This is inappropriate.

(f) Further, methods are frequently compared (see, for example, references 40 and 41) by comparing predictions made and announced *before* an experimental structure was known and secondary structural models made *after* a secondary structure is known (often referred to as "retrodictions") (42). The risks inherent in evaluating different tools for modelling protein conformation based on such comparisons have been noted elsewhere (43, 44), and we recommend that these be treated separately when evaluating prediction methods.

REFERENCES

- 1 Benner, S. A. (1989) *Adv. Enz. Reg.* 28, 219-236.
- 2 Benner, S. A. (1992) *Curr. Opin. Struct. Biol.* 2, 402-412.
- 3 Bazan, J. F. (1990) *Proc. Natl. Acad. Sci. USA* 87, 6934-6937.
- 4 Benner, S. A. and Gerloff, D. (1991) *Adv. Enz. Reg.* 31, 121-181.
- 5 Knighton, D. R., Zheng, J., Ten Eyck, L., Ashford, F. V. A., Xuong, N. H., Taylor, S. S. and Sowadski, J. M. (1991) *Science* 253, 407-414.
- 6 de Vos, A. M., Ultsch, M. and Kossiakoff, A. A. (1992) *Science* 255, 306-312.
- 7 Russell, R. B., Bredt, J. and Barton, G. J. (1992) *FEBS Lett.* 304, 15-20.
- 8 Pananoyotou, G., Bax, B., Gout, I., Federwisch, M., Wroblowski, B., Dhand, R., Fry, M. J., Blundell, T. L., Wollmer, A. and Waterfield, M. D. (1992) *EMBO J.* 11, 4261-4272.
- 9 Musacchio, A., Gibson, T., Lehto V.-P. and Saraste, M. (1992) *FEBS Lett.* 307, 55-61.
- 10 Benner, S. A., Cohen, M. A., and Gerloff, D. L. (1993) *J. Mol. Biol.* 229, 295-305.
- 11 Gerloff, D. L., Jenny, T. F., Knecht, L. J. and Benner, S. A. (1993) *Biochem. Biophys. Res. Comm.* 194, 560-565.
- 12 Gerloff, D. L., Jenny, T. F., Knecht, L. J., Gonnet, G. H. and Benner, S. A. (1993) *FEBS Lett.* 318, 118-124.
- 13 Chou, P. Y. and Fasman, G. D. (1978) *Adv. Enzymol.* 47, 45 - 148.
- 14 Garnier, J., Osguthorpe, D. J. and Robson, B. (1978) *J. Mol. Biol.* 120, 97-120.
- 15 Lenstra, J. A., Hofsteenge, J. and Beintema, J. J. (1977) *J. Mol. Biol.* 109, 185-193.
- 16 Maxfield, F. R. and Scheraga, H. A. (1979) *Biochemistry* 18, 697-704.
- 17 Nishikawa, K. and Ooi, T. (1986) *Biochem. Biophys. Acta* 871, 45-54.
- 18 Crawford, I. P., Niermann, T. and Kirschner, K. (1987) *Proteins: Struct. Funct. Gen.* 2, 118-129.

- 19 Taylor, W. R. and Geisow, M. J. (1987) *Prot. End.* 1, 183-187.
- 20 Zvelebil, M. J., Barton, G. J., Taylor W. R. and Sternberg, M. J. E. (1987) *J. Mol. Biol.* 195, 957-961.
- 21 Barton, G. J., Newman, R. H., Freemont, P. S. and Crumpton, M. J. (1991) *Eur. J. Biochem.* 198, 749-760.
- 22 Cohen, F. E. and Kuntz, I. D. (1987) *Proteins: Struct., Funct. Genet.* 2, 162-166.
- 23 Salzberg, S. and Cost, S. (1992) *J. Mol. Biol.* 227, 371-374.
- 24 Rost, B. and Sander, C. (1993) *J. Mol. Biol.* 231, 584-599.
- 25 Holley, H. L. and Karplus, M. (1989) *Proc. Nat. Acad. Sci. USA* 86, 152-156.
- 26 Kabsch, W. and Sander, C. (1983) *Biopolymers* 22, 2577-2637.
- 27 Bairoch, A. and Boeckmann, B. (1992) *Nucl. Acids Res.* 20, 2019-2022.
- 28 Dayhoff, M. O., Schwartz, R. M., and Orcutt, B. C. (1978) in *Atlas of Protein Sequence and Structure*. (Dayhoff, M. O., ed.), vol. 5, suppl. 3, pp. 345-352, National Biomedical Research Foundation, Washington, DC.
- 29 Pascarella, S. and Argos, P. (1992) *J. Mol. Biol.* 224, 461-471.
- 30 Benner, S. A., Cohen, M. A. and Gonnet, G. H. (1993) *J. Mol. Biol.* 229, 1065-1082.
- 31 Chothia, C. and Lesk, A. M. (1986) *EMBO J.* 5, 823-826.
- 32 Summers, N. L., Carlson, W. D. and Karplus, M. (1987) *J. Mol. Biol.* 196, 175-198.
- 33 Cohen, M. A., Benner, S. A. and Gonnet, G. H. (1994) *J. Mol. Biol.* submitted.
- 34 Kabsch, W. and Sander, C. (1983) *FEBS Lett.* 155, 179-182.
- 35 Robson, B. and Garnier, J. (1993) *Nature* 361, 506.
- 36 Colloch, N., Etchebest, C., Thoreau, E., Hennissat, B. and Moron, J.-P. (1993) *Protein Eng.* 6, 377-382.
- 37 Gomis-Rüth, F.-X., Kress, L. F. and Bode, W. (1993) *EMBO J.* 12, 4151-4157.
- 38 Zheng, D., Botos, I., Gomis-Rüth, F.-X., Doll, R., Blood, C., Njoroge, F. G., Fox, J. W., Bode, W. and Meyer, E. (1994) *Proc. Nat. Acad. Sci. USA.* submitted.
- 39 Thornton, J. M., Flores, T. P., Jones, D. T. and Swindells, M. B. (1991) *Nature* 354, 105-106.
- 40 Russell, R. B. and Barton, G. J. (1993) *J. Mol. Biol.* 234, 951-957.
- 41 Rost, B., Schneider, R. and Sander, C. (1993) *Trends Biochem. Sci.* 18, 120-123.
- 42 Toulmin, S. (1961) *Foresight and Understanding*, New York, Harper & Row, p 27.
- 43 Benner, S. A. (1994) *J. Molec. Recog.* in press.
- 44 Benner, S. A. and Gerloff, D. L. (1993) *FEBS Lett.* 318, 118-124.