PROTEINS: Structure, Function, and Genetics 23:446–453 (1995)

The Phospho-β-galactosidase and Synaptotagmin Predictions

Steven A. Benner, Dietlind Gerloff, and Gareth Chelvanayagam
Department of Chemistry, Swiss Federal Institute of Technology, CH-8092 Zürich, Switzerland

ABSTRACT **Two bona fide consensus predictions of secondary and tertiary structure in a protein family, made and announced before experimental structures were known, are evaluated in light of the subsequently determined experimental structures. The first, for phospho-β-galactosidase, identified the core strands of an 8-fold α–β barrel, and identified the 8-fold α–β barrel itself, which was found in the subsequently determined experimental structure to be the core folding domain. The second, for synaptotagmin, identified seven out of eight β-strands in the structure correctly, missing only a noncore strand. Three preferred "topologies" were selected from several hundred thousand possible topologies of these seven predicted strands using a rule-based analysis. The subsequently determined experimental structure showed that these seven strands in synaptotagmin adopt one of the three preferred topologies. We were unable, however, to identify the correct topology from among these three topologies.** © 1995 Wiley-Liss, Inc.

INTRODUCTION: SELECTION OF STRUCTURES

The ETH group prepared three predictions for the contest consummated in Asilomar. These were done in two phases. First, predictions for phospho-β-galactosidase[1,2] and isopenicillin N-synthase[3] were prepared. These proteins were chosen because they were large, mechanistically interesting enzymes with sufficient members in the evolutionary family to permit a consensus model to be built. In September, we were advised that the structure for isopenicillin N-synthase would not be solved in time for the contest, and asked to submit another prediction. At that time, only synaptotagmin (or C2 homology domain protein)[4,5] remained as a suitable target. Therefore, we prepared a prediction for synaptotagmin as well.[6] The synaptotagmin and phospho-β-galactosidase predictions are discussed below. An experimental structure for isopenicillin N-synthase is still not available.

SUMMARY OF THE PREDICTIONS

For both synaptotagmin and phospho-β-galactosidase, a residue-by-residue secondary structural model was first predicted.[2,6] Tertiary structural models were then built by assembling the predicted secondary structural elements. With phospho-β-galactosidase, the predicted secondary structure elements,[2] oriented using assignments of active site residues, suggested that the predicted secondary structural units assemble to form an 8-fold α–β barrel as the core domain.[2]

With synaptotagmin,[6] a combinatorial analysis was used to build a set of all possible β-sheet packing orientations for the seven β-strands that were predicted in the structure.[7] Following semiempirical rules, the number of possible folding "topologies" was reduced from several hundred thousand to just three. These corresponded to the β-sheet topologies found in the retinol binding protein, pseudoazurin, and the pleckstrin homology domain. Coordinate representations (see Figures 4, 5, and 6 in reference 6) of these three possible folds (or "topologies") were built based on the crystal structures of these domains or, in the case of the pleckstrin homology domain, on a model built from a published description of the structure.[6] Further, more speculative rules were then use to suggest that the pleckstrin homology domain model was best among the three preferred topologies.

EVALUATING THE PREDICTIONS

As a rule, we do not evaluate our own predictions until after they have been evaluated by others. Therefore, we are indebted to Thomas Defay and Fred Cohen for having served as judges in this contest.[8] Their insightful evaluation, appearing elsewhere in this issue, has enabled us to focus on a few points, important to a "right, wrong, why, learned" discussion of structure prediction generally, but illustrated well by these two predictions. Figures 1 (synaptotagmin) and Figure 2 (phospho-β-galactosidase) summarize the comparison between secondary structure as predicted and as assigned to crystallographic data.[9] These figures serve as reference for the discussion below.

Received April 18, 1995; revision accepted August 21, 1995.
Address reprint requests to Steven A. Benner, Department of Chemistry, ETH Zürich, CH-8092 Zürich, Switzerland.

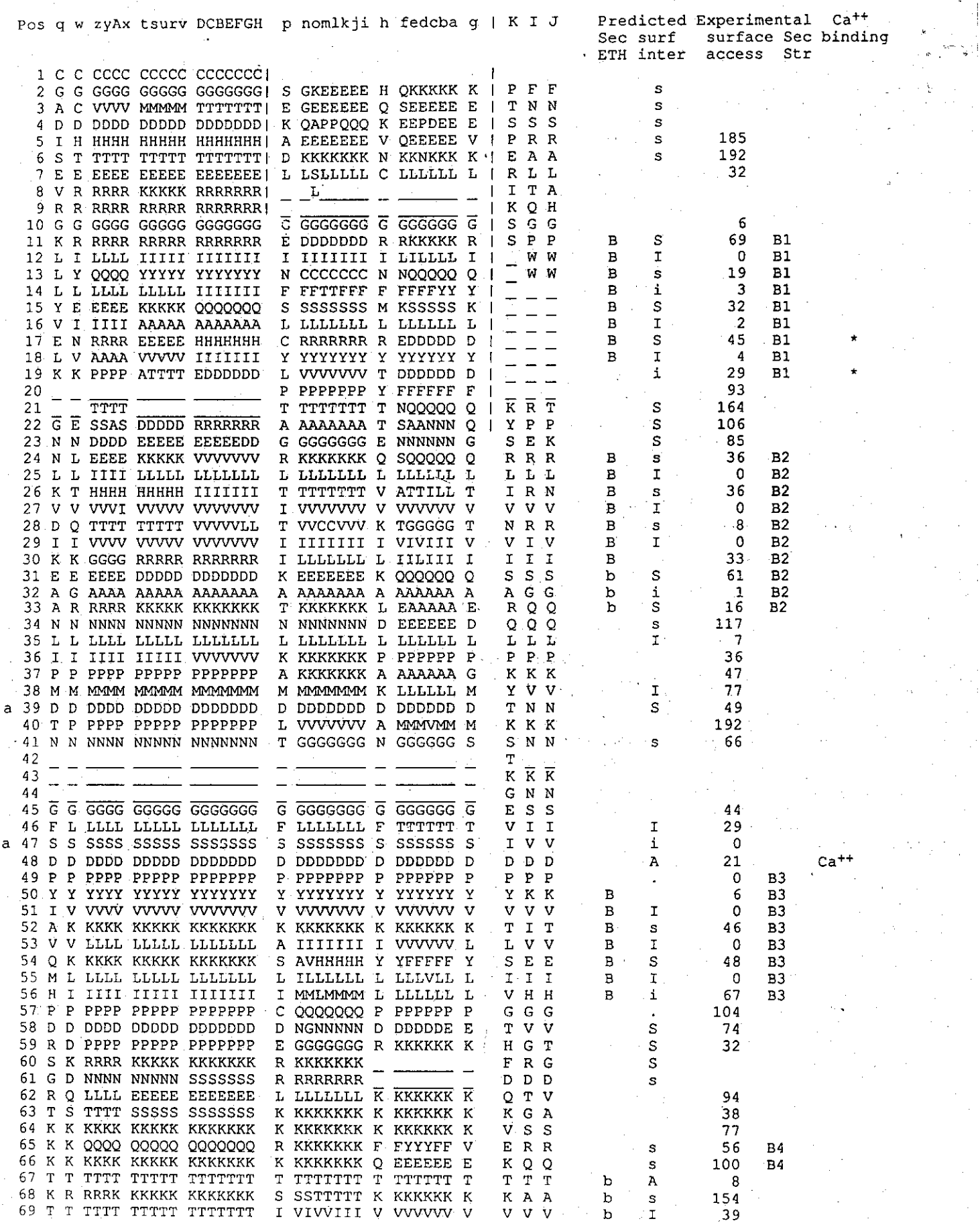

Fig. 1. Residue-by-residue secondary structure prediction for synaptotagmin. From left to right, the columns are alignment position number, multiple alignment using the one-letter code for amino acids, the predicted secondary structure, the predicted surface and interior assignments, the experimental surface and interior assignments, the experimental assignment of secondary structure, and the residues involved in calcium binding (asterisks refer to the "degenerate" calcium binding site.[9]) See Gerloff et al.[6] for details. (*Continued on overleaf.*)

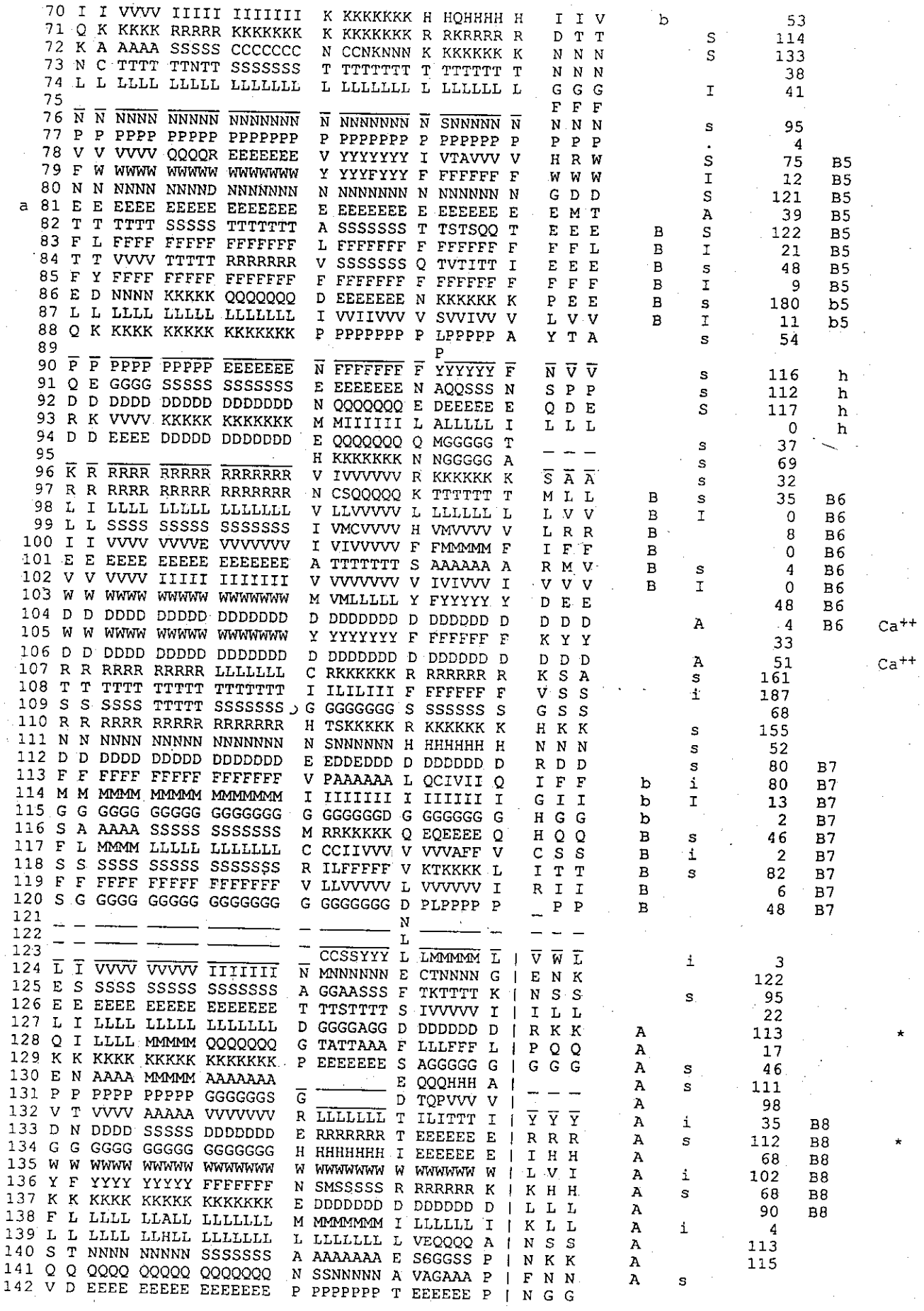

Fig. 1. (Continued from overleaf.)

```
MTKTLPKDFIFGGATAAYQAEGATHTDGKGPVAWDKYLEDNYWYTAEPAS 50
         eeeeeeeee                hhhhhhh           exp
        eee   eeeeee                              hh pred
        core strand 1

DFYHKYPVDLELAEEYGVNGIRISIAWSRIFPTGYGEVNEKGVEFYHKLF 100
hhhhhhhhhhhhhhh  eeeeeeeeeee           hhhhhhhhh exp
hhhhhhhhhhh   eeeee active site        hhhhhhhhhh pred
    core helix A   core strand 2       core helix B

AECHKRHVEPFVTLHHFDTPEALHSNGDFLNRENIEHFIDYAAFCFEEFP 150
hhhhhh eeeeee      hhhhhh       hhhhhhhhhhhhhhhhh  exp
hhhhh     eeeee                hhhhhhhhhhhhhhhhhhh pred
      core strand 3                   core helix C

EVNYWTTFNEIGPIGDGQYLVGKFPPGIKYDLAKVFQSHHNMMVSHARAV 200
  eeeee  hhhhhhhhhh             hhhhhhhhhhhhhhhhhh exp
h   eeee                         hhhhhhhhhhhhhhh   pred
core strand 4                      core helix D

KLYKDKGYKGEIGVVHALPTKYPYDPENPADVRAAELFDIIHNKFILDAT 250
hhhhh    eeeeeeeeeeee        hhhhhhhhhhhhhhhhhhhhh exp
           eeee               hhhhhhhhhhhhhh eeee pred
        core strand 5           core helix E

YLGHYSDKTMEGVNHILAENGGELDLRDEDFQALDAAKDLNDFLGINYYM 300
     hhhhhhhhhhhhhheeeeeee hhhhhhhhh    eeeeeeeee exp
      hhhhhhhh                    hhhhhh  eeeeee h pred
                                     core stand 6

SDWMQAFDGETEIIH.....SKYQIKGVGRRVAPDYVP.....WIIY 350
eeeeee    ee            eee    eeeee              exp
hhhhhhhhhh                                eeeee pred

PEGLYDQIMRVKNDYPNYKKIYITENGLGYKDEFVDNTVYDDGRIDYVKQ 400
 hhhhhhhhhhhh      eee   eee              hhhhhhhhh exp
 hhhhhhhhhhhhhh     eeee                  hhhhhhh pred
 core helix F     core strand 7              core

HLEVLSDAIADGANVKGYFIWSLMDVFSWSNGYEKRYGLFYVDFDTQERY 450
hhhhhhhhhh  eeeeeeeeee           eeee  eeeee  eeee exp
hhhhhhhhhh eeeee eeee                 eeeee       pred
helix G      core strand 8

PKKSAHWYKKLAETQVIE
e hhhhhhhhhhhh     exp
 hhhhhhhhhhhhh     pred
core helix H
```

Fig. 2. Experimental structure assigned by DSSP: h indicates an α-helix, e indicates a β-strand. The underlined regions designate the core secondary structural elements in the conserved α–β barrel domain. These are assigned using the DEFINE program; the differences between the DSSP and DEFINE assignments (made on the same experimental data) are worthy of note. This figure illustrates the accuracy of the consensus prediction in the assignment of secondary structure to elements of secondary structure that are conserved throughout the protein family, but not (by definition) to those that are not. See Gerloff and Benuer[2] for details.

USING TERTIARY STRUCTURAL INFORMATION (SURFACE AND INTERIOR ASSIGNMENTS) TO ASSIGN SECONDARY STRUCTURE: ARE THE SECONDARY STRUCTURE PREDICTIONS CORRECT FOR THE CORRECT REASONS?

The protein structure prediction problem embodies a "chicken-or-egg" paradox that arises because tertiary structural interactions are stronger than local sequence interactions in determining secondary structure.[10] This implies that predicting secondary structure from primary structure is essentially impossible without having at least some tertiary structural information.[11] At the same time, a reliable model for secondary structure appears to be necessary before predicting tertiary structure.

In these predictions,[12,13] this paradox is resolved by extracting tertiary structural information from a set of aligned homologous sequences before assigning a secondary structure.[14] This tertiary structural information comes primarily in the form of surface and interior assignments, although active site assignments are also obtained (see below). Patterns in these assignments are then used as indicators of secondary structure.

This approach has had reasonable success in many laboratories in making bona fide predictions of secondary structure,[15] those announced before an experimental structure is known. This success was mirrored in these predictions. For example, the synaptotagmin prediction identifies the first seven β-strands of the fold essentially correctly (Fig. 1). Further, with the exception of β4, there is excellent correspondence between the beginnings and ends of the strands as predicted and as experimentally assigned by DSSP. These include all of the strands that Sutton et al. identify as the "core" strands.[9]

It is important to ask whether the close correspondence between the predicted and experimental assignments of secondary structure in the synaptotagmin family was achieved for the correct reasons. Figure 1 shows that it was. Both the predicted assignment of secondary structure (S and s for strong and weak surface assignments, I and i for strong and weak interior assignment) and the experimental assignments (from DSSP) are given. For β1, β2, β3, β5, and β7, the correct assignment of secondary structure transparently arises from an accurate prediction of tertiary structural details, with the 2-residue periodicity of internal and external assignment characteristic of a β-strand.[16] Thus, the physicochemical basis for the secondary structure assignment in these segments is clear. β4 is too short to be analyzed in this fashion with statistical significance. The segment containing β6 was correctly identified as being largely internal, and the secondary structure correctly assigned using a different rule-based approach.

TRANSPARENCY IS USEFUL FOR IMPROVING A METHOD

There has been much discussion as to whether humans should be involved when predicting secondary structure, or whether a fully automated computer program should be the only goal of efforts in the field.[17] At one level, this discussion is purely technological: Does the man (woman) or the machine produce the better results? Defay and Cohen have addressed the issue in this context.[8]

We frequently pointed out that the issue can also be addressed in terms of research strategy.[11,12] Assume that we do not yet have a method that reliably assigns conformational information to a protein. Is it better at this point to design fully automated methods? Or to develop methods that involve human intervention? We have argued at some length that until the conformational problem is solved, human involvement in the bona fide prediction process is critical. Only through human involvement can one understand why a method works (when it works) and why it fails when it fails. It is this understanding that underlies rational improvement of prediction methods. Thus, a transparent method that performs poorly can often be more useful for developing the underlying science than a nontransparent method, even if it performs better.

For example, the synaptotagmin prediction is marred by a serious error: the final strand was misassigned as a helix (Fig. 1). With a transparent prediction tool, we are able to ask why. Inspection of the multiple alignment in Figure 1 shows that two factors figured significantly in the misassignment. First, a manual readjustment of the multiple alignment shifted a gap from the middle of the region mispredicted as a helix. As a gap is regarded as a "parsing" element, this shifting was necessary to predict the helix; had the gap not been shifted, a helix would not have been assigned in this region. Here, human intervention diminished the technological quality of the prediction. However, the transparency of the process allows us to understand what happened that led to the misassignment. In particular, the fact that the alignment is difficult to construct in this region suggests that secondary structure may have diverged in this region. This leads to further analysis below.

Second, the misassignment was aggravated because of an incorrect assignment of a conserved tryptophan at position 135 to the inside. This Trp residue lies on the surface in the experimental structure. This misassignment obscured a tertiary structural assignment pattern that would have identified this segment as a strand. It is interesting to note that in the broader multiple alignment provided by Sutton et al.,[9] this Trp is not conserved, but rather displays a pattern of variation and conservation characteristic of a surface residue (but see below concerning the quality of the multiple alignment). A similar problem was encountered in the prediction of the SH3 domain,[18] and future heuristics for assigning surface/interior residues will flag conserved tryptophan residues.

WHAT DOES IT MEAN WHEN ONE SAYS THAT "WHEN THE QUALITY OF AN ALIGNMENT IS POOR, PREDICTION ACCURACY SUFFERS"?

It is frequently noted that predictions based on multiple alignments suffer when the multiple alignment is poor. This comment creates the impression that if only we could improve tools for creating multiple alignments, than we would be better able to predict secondary structure. This impression is mis-

taken, in a way that is important for the future development of prediction tools.

A structural prediction made for a family of aligned homologous protein sequences is a consensus prediction. It is based on the assumption that proteins related by common ancestry all have similar conformation. This assumption is true only as an approximation. As proteins within a family diverge at the level of sequence, they also diverge at the level of conformation, first in loops, then in side chain orientations, then in secondary structure, and finally through the modification of conformation of entire domains. An excellent quantitative statement of this divergence was given by Chothia and Lesk.[19]

The breakdown of the assumption underlying all evolution-based structure prediction methods has an impact on the way in which predictions are evaluated.[20,21] A consensus model can be accurate for only those structural features that a protein family holds in common.[21] Further, as a consensus model is generally evaluated by examining only a single member of the protein family that it covers, at least some of the "mistakes" reflect not errors introduced by the prediction method, but rather cases where the individual experimental structure has a conformation different from the conformation of other proteins in the family. This represents, of course, an intrinsic limitation of the consensus modeling strategy.

Regions where the conformation has diverged within a family are, as a rule, regions that are difficult to align. Conversely, regions that are difficult to align are, quite often, regions where conformation has diverged. Regions that are difficult to align are therefore regions where the assumption underlying all evolution-based structure prediction tools is likely not to apply. Thus, these are regions where a consensus prediction is likely to be judged to be "wrong," especially when a single experimental structure is used to evaluate the prediction, not because the tool used to make the prediction was flawed, but because the strategy itself is based on an assumption that is not universal.

Fortunately, secondary structural elements that are not conserved rarely lie at the core of the folded structure. Thus, as noted by Defay and Cohen,[8] the nonuniversality of the assumption does remarkably little damage when building tertiary structural models from a predicted consensus secondary structural model. For example, the phospho-β-galactosidase family contains two major subfamilies with low sequence similarity overall.[2] Large segments of the protein are difficult to align, implying that in these regions, secondary structure has diverged substantially. As a result, the single known experimental structure, obtained for protein "a" (Subfamily a) in the alignment,[2] is insufficient to evaluate the secondary structure in regions where secondary structure has widely diverged (Fig. 2); several of the elements predicted in the consensus model are simply wrong when evaluated using this particular experimental structure.[8] Nevertheless, because core structural regions are conserved in both subfamilies of the phospho-β-galactosidase family and because these are the important elements when assembling a tertiary structural model, the consensus prediction (Fig. 2) is adequate as a starting point to build an essentially correct tertiary structural model for the "core" fold, the fold of the protein that both subfamilies adopt in common.

Likewise, the misassigned helix in the synaptotagmin family lies in a region where the alignment is extremely poor. This is, of course, the reason why the gap was misplaced in this region, and one reason why the secondary structural element was misassigned (see above). However, the information that we presently have does not allow us to conclude that a strand should be predicted in this region in the consensus model, as secondary structure within this region may not be conserved. We must wait for more experimental structures from the synaptotagmin superfamily before we will know whether this misassignment reflects a weakness in the secondary structure prediction heuristic, or whether it reflects divergence in secondary structure within the family, and therefore a weakness of the consensus modeling approach generally. Again, however, the misassignment of a noncore secondary structural element did not obstruct the assembly of the correct tertiary structure for synaptotagmin as one of three alternative packings (see below).

TERTIARY STRUCTURAL MODELING VERSUS THREADING

Given reliable tools for assigning secondary structure, the task remaining is to build a tertiary structural model. One approach is to attempt to "thread" a predicted secondary structural model on to proteins known in the database.[22] Alternatively, one might attempt to assemble the predicted secondary structural elements de novo.

In the phospho-β-galactosidase prediction, the secondary structure pattern was obviously a signature of an 8-fold α–β barrel.[23] Further, the placement of active site residues in the model confirmed the β-barrel as the preferred topology for the overall fold. As the tertiary fold was identified in the context of a large number of known proteins with similar folds, the tertiary structure can be said to have been predicted by a "threading" approach in combination with a de novo secondary structure prediction. In this respect, this prediction is not distinctly different from the identification of a similar fold by a similar method by Kirschner and co-workers 7 years ago using a consensus secondary structure prediction obtained by the GOR method.[24,25] As the reliability of secondary structure prediction tools

based on multiple sequence alignments has increased substantially since then, this sort of pattern recognition exercise will be routine in the future.

For synaptotagmin, the prediction of tertiary structure was more complicated. The secondary structure prediction did not in itself indicate a particular topology, only that the core structure was an all-β structure. There are, of course, many classes of β-sheets, barrels, and sandwiches, and neither the secondary structure prediction nor the active site assignments identified one of these as a clear template for building a tertiary prediction. Therefore, tertiary structural models were assembled de novo.[6] First, a combinatorial approach was used to assemble all 322,560 possible sheet structures from the predicted β-strands. Then, a large majority of these were excluded by enforcing connectivity of strands in a β-sheet, avoiding loop crossovers, and using other rules that have (at least some) empirical basis.[26] This process reduced the number of possible β-sheet topologies first to 36, and then to six, grouped as three pairs of alternative folds. The details of this analysis are found in Gerloff et al.[6]

The database was then examined for analogs for the three folds remaining. The first, where the strands were placed consecutively in an ABCDEFG "up-down" pattern, found its closest analog in the retinol binding protein.[27] Including a single Greek key element in the fold approximated the fold found in pseudoazurin (ABEDCFG).[28] To make the analogy to pseudoazurin "work," the first strand of pseudoazurin was ignored, and a strand was moved from one sheet in the β-sandwich to the other (the "modified pseudoazurin fold").[6] The third alternative fold (ABCDGFE) had a topology similar to that found in the pleckstrin homology.[29] The "modified pseudoazurin" fold (ABEDCFG) turned out to be the correct topology for the fold of first seven β-strands in synaptotagmin.[9]

The fact that the complexity of the tertiary structure modeling problem was reduced in this case from hundreds of thousands of possible topologies to just three, with the correct topology contained within the three, makes several points. First, the misassignment of the final strand in the fold as a helix did not interfere with the assignment of the tertiary fold. This was undoubtedly due to the fact that the misassigned secondary structural unit was not at the core of the fold (see above), and also because it came at the end of the domain, not at the middle. Of course, the fact that the segment was misassigned and that it is not a core segment is not independent, as noted above.

Second, we were unable to identify the correct tertiary fold from among these three alternatives. Efforts were made to identify hypotheses to try to create a preference for one of the three folding topologies. We noted that these hypotheses were little more than speculative efforts to use "a bona fide prediction opportunity to test some unorthodox ideas." Nevertheless, it is important to note that these unorthodox ideas failed. We preferred the pleckstrin homology domain fold over the pseudoazurin fold; as noted above, synaptotagmin in fact adopts (in its first 7 strands) a modified pseudoazurin topology. Application of a threading program might have been a better approach for identifying the correct fold from among the three preferred topologies.[30]

Another noteworthy aspect of the synaptotagmin prediction is its model built for the calcium-binding active site. In the prediction, Asp-48 (Asp-178 in the synaptotagmin numbering), Asp-104 (Asp-230), Asp-106 (Asp-232), and Glu-81 (Glu-208) were predicted to form a calcium-binding site. In making this prediction, only patterns of conservation and variation within the multiple alignment were considered, together with the knowledge that synaptotagmin most likely contained a calcium-binding site. This model proved to be a good representation for the putative calcium binding active site in synaptotagmin, which is built from residues Asp-48, Asp-104, and Asp-106 (synaptotagmin numbering 178, 230, and 232). Thus, the active site assignments made in the prediction were useful guides for predicting tertiary structure, as they were in the phospho-β-galactosidase prediction.

WHAT HAVE WE LEARNED?

These two predictions add to the dozen or so remarkably accurate bona fide predictions that have now been made in many laboratories using aligned homologous protein sequences.[15] As the collection of examples builds, much is learned about both the specific heuristics used to assign structural features of a protein, but also about protein folding in general. It seems that it is possible to obtain reasonably reliable predictions of surface and interior assignments from a set of aligned homologous protein sequences, and to use this tertiary structural information, together with a rule-based approach based for internal segments to obtain a plausible model of secondary structure, at least for core elements. These have been sufficient, at least in these cases, to narrow the number of possible folding topologies to a small number. In combination with threading tools, the number of tertiary models for the core can be reduced to one (for phospho-β-galactosidase). In other cases, additional tools, not yet developed, are needed to identify a unique prediction from a small set of alternative structures. We look forward to the next contest 2 years hence.

REFERENCES

1. Hengstenberg, W., Kohlbrecher, D., Witt, E., Kruse, R., Christiansen, I., Peters, D., Vonstrandmann, R.P., Stadtler, P., Koch, B. Structure and function of proteins of the phosphotransferase system and of 6-phospho-beta-ga-

lactosidases in Gram-positive bacteria. FEMS Microbiol. Rev. 12:149–164, 1993.
2. Gerloff, D.L., Benner, S.A. A consensus prediction of the secondary structure for the 6-phospho-β-D-galactosidase superfamily. Proteins 21:273–281, 1995.
3. Benner, S.A., Jenny, T.F., Cohen, M.A., Gonnet, G.H. Predicting the conformation of proteins from sequences. Progress and future progress. Adv. Enzyme Reg. 34:269–353, 1994.
4. Perin, M.S., Fried, V.A., Mignery, G.A., Jahn, R., Suedhof, T.C. Phospholipid binding by a synaptic vesicle protein homologous to the regulatory region of protein kinase C. Nature (London) 345:260–263, 1990.
5. Pevsner, A., Scheller, R. H. Mechanisms of vesicle docking and fusion. Insights from the nervous system. Curr. Opin. Cell Biol. 6:555–560, 1994.
6. Gerloff, D.L., Chelvanayagam, G., Benner, S.A. A predicted consensus structure for the protein kinase C2 homology (C2H) domain, the repeating unit of synaptotagmin. Proteins 22:299–310, 1995.
7. Cohen, F.E., Abarbanel, R.M., Kuntz, I.D., Fletterick, R. J. Secondary structure assignment for alpha-beta-proteins by a combinatorial approach. Biochemistry 22:4894–4904, 1983.
8. Defay, T., Cohen, F.E. Evaluation of current techniques for ab initio protein structure prediction. Proteins 23:431–447, 1995.
9. Sutton, R.B., Davletov, B.A., Berghuis, A.M., Südhof, T.C., Sprang, S.R. Structure of the first C2 domain of synaptotagmin I: A novel Ca^{2+} phospholipid binding fold. Cell 80:929–938, 1995.
10. Cohen, B.I., Presnell, S.R., Cohen, F.E. Origins of structural diversity within sequentially identical hexapeptides. Protein. Sci. 2:2134–2145, 1993.
11. Benner, S.A. Patterns of divergence in homologous proteins as indicators of tertiary and quaternary structure. Adv. Enzyme Reg. 31:121–181, 1989.
12. Benner, S.A., Gerloff, D. Patterns of divergence in homologous proteins as indicators of secondary and tertiary structure: The catalytic domain of protein kinases. Adv. Enzyme Reg. 31:121–181, 1991.
13. Benner, S.A. Predicting de novo the folded structure of proteins. Curr. Opin. Struct. Biol. 2:402–412, 1992.
14. Benner, S.A., Badcoe, I., Cohen, M.A., Gerloff, D.L. Bona fide prediction of aspects of protein conformation. Assigning interior and surface residues from patterns of variation and conservation in homologous protein sequences. J. Mol. Biol. 235:926–958, 1994.
15. Benner, S.A., Gerloff, D.L., Jenny, T.F. Predicting protein crystal structures. Science 265:1642–1644, 1994.
16. Lim, V.I. Algorithms for prediction α-helices and β-structural regions in globular proteins. J. Mol. Biol. 88:873–894, 1974.
17. Gerloff, D.L., Benner, S.A. Predicting the conformation of proteins: Man versus machine. FEBS Lett. 325:29–33, 1993.
18. Benner, S.A., Cohen, M.A., Gerloff, D.L. A predicted secondary structure for the Src homology domain 3. J. Mol. Biol. 229:295–305, 1993.
19. Chothia, C., Lesk, A. The relation between the divergence of sequence and structure in proteins. EMBO J. 5:823–826, 1986.
20. Zhu, Z.Y. A new approach to the evaluation of protein secondary structure predictions at the level of the elements of secondary structure. Prot. Eng. 8:103–108, 1995.
21. Jenny, T.F., Benner, S.A. Evaluating predictions of secondary structure in proteins. Biochem. Biophys. Res. Commun. 200:149–155, 1994.
22. Bryant, S.H., Lawrence, C.E. An empirical energy function for threading protein, sequence through the folding motif proteins. 16:92–112, 1993.
23. Farber G.K., Petsko, G. The evolution of α/β barrel enzymes. Trends Biochem. Sci. 15:228–234, 1990.
24. Crawford, I.P., Niermann, T., Kirschner, K. Prediction of secondary structure by evolutionary comparison: Application to the α subunit of tryptophan synthase. Protein 2:118–129, 1987.
25. Garnier, J., Robson, B. The GOR method for predicting secondary structures in proteins. In: "Prediction of Protein Structure and the Principles of Protein Conformation." G.D. Fasman, (ed.). Plenum, New York: 1989: 417–465.
26. Woolfson, D.N., Evans, P.A., Hutchinson, E.G., Thornton, J.M. Topological and stereochemical restrictions in beta-sandwich protein structures. Prot. Eng. 6:461–470, 1993.
27. Zanotti, G., Berni, R., Monaco, H.L. Crystal structure of liganded and unliganded forms of bovine plasma retinol binding protein. J. Biol. Chem. 268:10727–10738, 1993.
28. Petratos, K., Banner, D.W., Beppu, T., Wilson, K.S., Tsernoglou, D. The crystal structure of pseudoazurin from Alkaligenes faecalis SS-6 determined at 2.9 Å resolution. FEBS Lett. 218:209–214, 1987.
29. Yoon, H.S., Hajduk, P.J., Petros, A.M., Olejniczak, E.T., Meadows, R.P., Fesik, S.W. Solution structure of a pleckstrin-homology domain. Nature (London) 369:672–675, 1994.
30. Jones, D.T., Taylor, W.R., Thornton, J.M.A new approach to protein fold recognition. Nature (London) 358:86–89, 1992.