

The Anatomy & Taxonomy of Protein Structure

by Jane S. Richardson

IV. Discussion

A. Implications for Noncrystallographic Determination of Protein Structure

1. Principles and Methods

Aside from the direct techniques of X-ray or electron diffraction, the major possible routes to knowledge of three-dimensional protein structure are prediction from the amino acid sequence and analysis of spectroscopic measurements such as circular dichroism, laser Raman spectroscopy, and nuclear magnetic resonance. With the large data base now available of known three-dimensional protein structures, all of these approaches are making considerable progress, and it seems possible that within a few years some combination of noncrystallographic techniques may be capable of correctly determining new protein structures. Because the problem is inherently quite difficult, it will undoubtedly be essential to make the best possible use of all hints available from the known structures.

The most important general point to be emphasized is that it is now both possible and essential to frame and test hypotheses about exactly which structural features a given technique is really measuring or predicting. Structure surveys like the current one can help in choosing proteins that will provide critical tests of such hypotheses both by locating proteins that vary most in the parameters under consideration, and also in helping to control for the effects of differences in other major structural parameters. It seems inherently unlikely that any spectroscopic feature is a direct measure of percentage α , β turn, and coil as defined in any of the usual ways. But since such percentages are certainly not the only useful way to describe protein structure, it should be fruitful to combine theoretical analysis (where possible) with careful empirical tests in order to determine the set of descriptions most applicable for a given technique. Let us consider several examples of what could be attempted with this approach.

Methods for predicting secondary structure from amino acid sequence could presumably benefit from considering parallel and antiparallel β sheet separately, since the two types have rather different single and pairwise residue preferences. An overall classification scheme could help in

choosing a large and characteristic sample. However, there is the difficulty of dealing with mixed β sheets. For a given set of parameters that successfully distinguished pure parallel from pure antiparallel sheets, it would be possible for instance to test whether the characteristics of strands in mixed β sheets depended mainly on their local hydrogen-bonding type or depended mainly on whether the overall sheet organization was "antiparallel β " type or "parallel α/β " type (for example, prealbumin is an "antiparallel" mixed sheet and carboxypeptidase is a "parallel" mixed sheet). This sort of question should be asked also for the amide I bands in infrared (see Miyazawa and Blout, 1961; Chirgadze and Nevskaya, 1976) and Raman spectra (see Krimm and Abe, 1972; Yu et al., 1975) that are thought to be sensitive to the differences between parallel and antiparallel β sheet. It would be especially useful if it turned out that some features were sensitive to local and some to overall structure. In general, the parallel α/β structures have been grossly underrepresented in spectroscopic studies of protein conformation, because they do not occur in the small proteins that made up most of the early X-ray structure determinations. Now that α/β proteins have been shown to be extremely common, this sampling bias can be corrected.

The C–S and S–S stretch vibrations of disulfides (Edsall et al., 1950) can be observed in the Raman spectra of proteins, but their interpretation is still somewhat controversial (see, for example, Klis and Siemion, 1978; Spiro and Gaber, 1977). Using series of model compounds, Van Wart et al. (1973) have related S–S stretch frequency to the χ_3 (C β –S–S–C β) dihedral angle, while Sugeta et al. (1972, 1973) have related the S–S frequency to the χ_2 (C α –C β –S–S) dihedral angle and C–S stretch frequency to χ_1 angle; these latter correlations have been further modified by Van Wart and Scheraga (1976). The relationship of spectrum to conformation seems to be quite complex in proteins, where constraints at either end of the disulfide would tend to increase coupling between the modes. The S–S stretch may be sensitive to the relative sign as well as the absolute value

of χ_2 , and therefore may reflect the difference between the spiral and the hook conformations (see Section II,E). It should be possible to determine characteristic spectra for the three common disulfide conformations found in proteins (the left-handed spiral, the right-handed hook, and the extended form in immunoglobulins) by choosing accurately refined proteins with a single or a dominant disulfide conformation (e.g., immunoglobulins, carboxypeptidase, egg lysozyme, and pancreatic trypsin inhibitor).

Very low-frequency vibrations have been observed in proteins (e.g., Brown et al. 1972; Genzel et al., 1976), which must involve concerted motion of rather large portions of the structure. By choosing a suitable set of proteins to measure (preferably in solution), it should be possible to decide approximately what structural modes are involved. Candidates include helix torsion, coupled changes of peptide orientation in β strands, and perhaps relative motions of entire domains or subunits. These hypotheses should be tested, because the low frequency vibrations probably reflect large-scale structural properties that would be very useful to know.

In using circular dichroism to estimate percentages of the various secondary structures in a protein (e.g., Saxena and Wetlaufer, 1971; Grosse et al., 1974), helix can be judged more reliably than other features, as is usually true for almost any method including prediction (e.g., Maxfield and Scheraga, 1976). This is presumably because α -helices are relatively uniform in both local and longer range patterns, while β structure is widely variable in hydrogen-bonding pattern, regularity, twist, exposure, and overall shape. There is at least a real possibility that differences in shape and organization of β structure are reflected in the circular dichroism spectrum; that possibility should be tested, because it would be even more useful to be able to categorize a structure as a doubly wound sheet or an antiparallel β barrel than to say it had 35% β structure, even supposing that we could reliably do the latter.

Successful examples of the sort of correlations postulated above would add additional independent pieces of information for use in a combined strategy of noncrystallographic protein structure determination. Empirical regularities such as the handedness of crossover connections (see Section II,B) can help in narrowing down the possibilities. Another need is to decide whether, and at what point, a protein is divided into domains. The more tenuously connected domain pairs can often be recognized by such techniques as electron microscopy, viscosity, low-angle scattering, or proteolysis, and it might prove possible to recognize domain-connection regions in the sequence.

Knowledge of a set of common overall structure types (such as the major subgroupings in our classification scheme) can provide prototypes with which to match the distribution of predicted secondary structures and the characteristics suggested by various spectroscopic measures. For a given protein, combination of all these methods in an overall strategy that can deal with their probabilistic nature and disparate information content may some day be able to produce a fairly small number of alternative structures, one of which (by some process such as energy minimization) would converge to what could be recognized as the correct native structure.

Even an infallible method of structure prediction would not make protein crystallography obsolete; detailed prior knowledge of the globin structure has not removed the necessity or interest of high resolution X-ray structures for other species, mutants, and ligand forms of hemoglobin. What it would do is to take away a great deal of the fun and excitement of discovering new structures by protein crystallography; but that is not too large a price for the kind of increased understanding that is likely to accompany even the most ad hoc of successful structure prediction methods.