SOME METHODS OF REPRESENTATION OF PROTEIN CRYSTALLOGRAPHIC STRUCTURAL DATA

R. SRINIVASAN AND R. BALASUBRAMANIAN

Centre of Advanced Study in Physics, University of Madras, Guindy Campus, Madras 600 025, India

THE data in the form of three-dimensional atomic coordinates that results from an X-ray analysis of a protein crystal form the basis for extracting different types of information about the tertiary structure of a protein molecule. Conventional method of model building modified suitably as in the Kendrew models give an overall picture of the intricate three-dimensional architecture. The complete specification of a molecule by the three-dimensional Cartesian atomic coordinates may be replaced in principle by a different set of parameters such as the conformational angle parameters. Thus for instance the set of ϕ–ψ angles[1] at each C^α-atom specifies completely the relative positions of the backbone atoms in the protein chain. Other conformational parameters are also needed[2] to make the specification of a molecule complete, including the side chains. The use of the ϕ–ψ parameters is convenient for the description of the chain folding and is used widely, in particular, in the various approaches to prediction of protein conformations by theoretical and semi-empirical methods[3]. The three-dimensional models are at times felt to be unwieldy. The ϕ–ψ diagram although is simple and powerful, is not convenient for certain purposes, such as when one wishes to find which parts of the molecule are close to each other and so on. The so-called distance map[4,5] can yield this information and has been found more useful in drawing certain conclusions on the evolutionary correlations[6] in proteins through similarities in their patterns. Other types of analyses have also been reported for other purposes such as plotting of the distance ($l_{i,\,i+3}$) between C^α atoms separated by three residues[7] or the number of C^α-atoms within a specified distance from a given C^α atom[5]. Recently we have examined quite generally different types of such representation of the three-dimensional data that are possible. This is a short report of some of the main findings from our investigations. Broadly the different types of analyses may be classified into two categories of representation of the available data, namely, one- and two-parameter representations.* The ϕ–ψ diagram and the $(i-j)$ distance map may be considered as two two-parameter representations. The chain-plot† of $l_{i,\,i+3}$ is typically a one-parameter representation.

A few other possibilities we have tried are as follows: firstly, we have tried the torsion angle involving the four atoms C^α_{i-1}, C^α_i, C^α_{i+1} and C^α_{i+2} around the virtual bond $C^\alpha_i - C^\alpha_{i+1}$ denoted by θ_i, for a chain-plot. Figure 1 shows a typical chain-plot of θ_i for myoglobin. Unlike the $l_{i,i+3}$-plot the θ values have a wide range, namely, $-180°$ to $+180°$ and in addition they also give information about the sense of folding of the chains as one progresses along the chain. Standard conformations such as the α-helix have characteristic values of θ which can be calculated from the standard dimensions and ϕ, ψ values of a pair of peptide units. For the α-helix θ turns out to be 50°. The α-helical regions are seen to have fairly a constant value around 50° (Fig. 1).

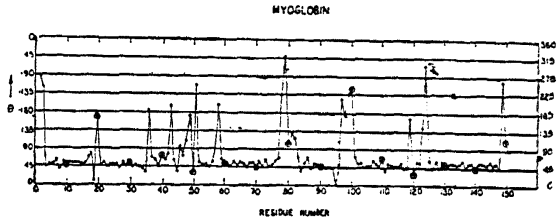

FIG. 1. Chain-plot of θ_i for myoglobin. For convenience every tenth residue is marked by.

During these investigations we have found it convenient to evolve a few new parameters which enable us to characterise helical segments and their analysis. Some of the one- and two-parameter representations to be described below involve these new parameters.

Thus considering the C^α atoms of a chain to lie on a perfect helix (*i.e.*, with $C^\alpha_i - C^\alpha_{i+1}$ lengths and $\delta_i = C^\alpha_{i-1} - C^\alpha_i - C^\alpha_{i+1}$ angles all equal) a consecutive group of four atoms, C^α_{i-1}, C^α_i, C^α_{i+1}, C^α_{i+2} may be considered to form the smallest segment of the helix. If $\boldsymbol{h}_i$ denotes the unit vector along the direction of the axis for the above segment, one may consider in general the directions of the axes of the different helical segments given by $\boldsymbol{h}_i$, $\boldsymbol{h}_{i+1}$ etc., and devise parameters for representations. For

* In principle one should include the possible three-parameter representations also. We shall not consider these since the source itself is such a three-parameter representation and our aim is at simplified analysis.

† We use the term chain-plot to denote plotting of the value of any chosen parameter as a function of the residue number in the chain.

instance, the angle between the axes‡ of successive segments i and $i-1$, denoted by η_i, may be used for a chain-plot. In regions of perfect helix these values will have a constant zero value signifying perfect alignment of successive segments. A typical η_i plot for myoglobin is shown in Fig. 2.

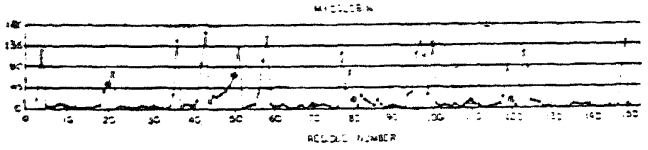

FIG. 2. Chain-plot of η_i for myoglobin. For convenience very tenth residue is marked by.

With the available axial directions h_i, h_j one may also plot the interaxial angles η_{ij} for two segments i and j and is thus a two-parameter representation. This is somewhat similar to the $l_{i,j}$ distance map excepting that the parameter used now is an angle. Figure 3 shows the η_{ij} map for myoglobin. α-helixal regions are characterised by triangular blocks along the diagonal.

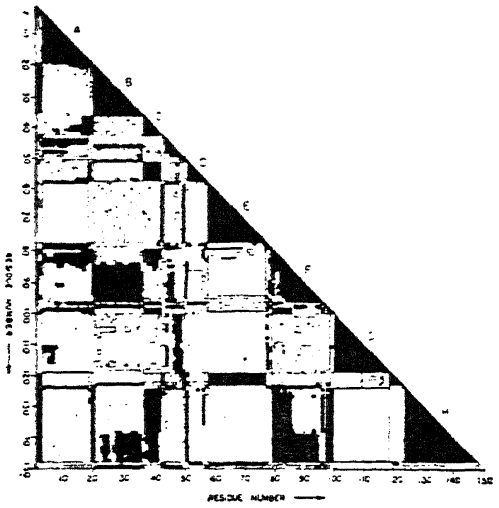

FIG. 3. η_{ij}-plot for myoglobin. The various helical segments A, B, etc., are marked along the diagonal. Dark, gray and blank regions correspond respectively to η ranges of 0° to 60°, 60° to 120° and 120° to 180°.

The other representation may be described more appropriately as a two-dimensional representation. Here one makes use of the axial direction h_i for the various segments and plots a stereographic projection, well known to crystallographers. This projection has the advantage that the relative angles of directions come out conveniently while the information about distances in the structure completely disappear. Thus for instance the relative angular orientation of say α-helical segments can be readily discerned (Fig. 4). Ideally for a

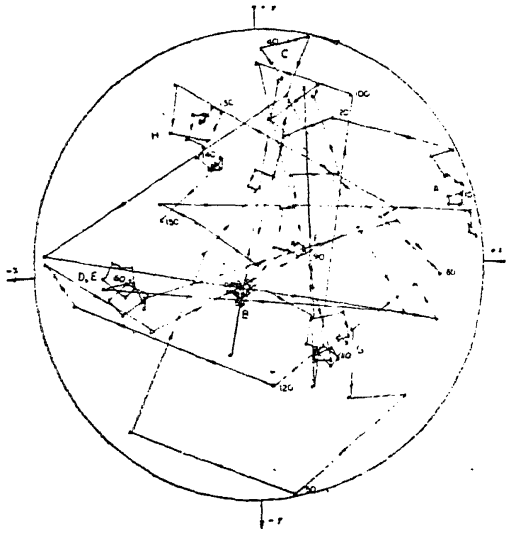

FIG. 4. Stereographic projection for h_i in myoglobin with $+Z$ pointing upwards. Helical regions A, B, etc., are marked near respective segments.

perfect helix the projection of points corresponding to successive segments should superpose in projection but in practice one obtains only a dense distribution about a point corresponding to the mean direction. The density of packing of the points also indicates the relative tightness or perfection of the helices (*e.g.*, compare helix E and B in Fig. 4).

Further investigations are in progress and will be reported in due course.

ACKNOWLEDGEMENT

The authors wish to thank Mr. S. S. Rajan for helping them in some of the calculations.

1. Ramachandran, G. N., Ramakrishnan, C. and Sasisekharan, V., *J. Mol. Biol.*, 1963, 7, 95.
2. IUPAC-IUB Commission on Biochemical Nomenclature, *Biochemistry*, 1970, 9, 34, 71.
3. Burges, A. W., Ponnuswamy, P. K. and Scheraga, H. A., *Israel J. Chem.*, 1974, 12, 239.
4. Phillips, D. C., In : *British Biochemistry, Past and Present* (Goodwin, T. W., Ed).), Academic Press, London, 1970, p. 11.
5. Ooi, T. and Nishikawa, K., In : *The Jerusalem Symposia on Quantum Chemistry and Biochemistry*, (Bergmann, E. D. and Pullman, B., Eds.), Jerusalem Academic Press, 1973, 5, 173.
6. Rossman, M. G. and Liljas, A., *J. Mol. Biol.*, 1974, 85, 177.
7. Watson, H. C., *Progress in Stereochemistry*, 1969, 4, 229.

‡ In an actual case of a protein, a segment of four atoms C^{α}_{i-1}, C^{α}_{i}, C^{α}_{i-1}, C^{α}_{i-2} need not form a part of a regular helix. Still a helical axis direction $\mathbf{h}_i$ for such a segment can be found as a first approximation and used for our purposes.