Can multiple low-density immunoadsorbents and ‘negative’ distance matrices help predict the three-dimensional structure of any globular protein?

Shiv PILLAI

Kothari Centre of Gastroenterology, The Calcutta Medical Research Institute, 7/2 Diamond Harbour Road, Calcutta-700 027, India

(Received 29 June 1981/Accepted 6 July 1981)

Chemical cross-linking of a protein immobilized on multiple, site-specific, monoclonal, chemically modified antibody matrices combined with analysis of data on a distance matrix is likely to yield sufficient positive and ‘negative’ long-range distance information to help predict the three-dimensional structure of any single-chain globular protein.

The understanding of any biological function or event in molecular terms depends on knowledge of the three-dimensional structures of the macromolecules involved. Although the X-ray-diffraction analysis of protein crystals is undoubtedly the most precise means of establishing the structure of a protein, the realization that a vast number of ‘interesting’ proteins may, perhaps, never prove crystallizable has catalysed the growth of methods aiming to predict the conformation of a protein from its amino acid sequence (Anfinsen & Scheraga, 1975; Guzzo, 1965; Schulz *et al.*, 1974; Matthews, 1975). The underlying principle of all these predictive approaches is the determination of the folded state that possesses the least free energy; for globular proteins, however, these approaches have been consistently hampered by the ‘multiple-minimum’ problem (the existence of many local minima in the multidimensional energy surface), a problem for which no clear solution appears to be in sight (Swenson *et al.*, 1978). Even the best-case expectations for non-crystallographic approaches such as conventional cross-linking methods or n.m.r. techniques fall far short of the requirements for restricting structures to an acceptably small region of conformation space (Havel *et al.*, 1972).

Conventional cross-linking methods have been most useful in studying ‘nearest-neighbour’ interactions in subcellular organelles (Bickle *et al.*, 1972; Steck, 1972; Thomas & Kornberg, 1975; Tinberg *et al.*, 1975). Although useful information regarding inter-residue distances in single-chain globular proteins (Hartmann & Wold, 1967) and the subunit stoichiometry (Davies & Stark, 1970) and orientation of protomers (D’Albis & Gratzer, 1976) in oligomeric proteins has occasionally been obtained,

the information from ‘in-solution’ cross-linking methods regarding the tertiary structure of proteins is too limited to be of significant value in predictive studies of protein topology.

The major limitation (Peters & Richards, 1977) of all previous cross-linking approaches for the study of protein structure has been their necessarily low efficiency (efficient procedures would favour intermolecular conjugation) and consequent inability to provide any reliable ‘negative’ data, that is, unambiguous structural interpretations made on the basis of cross-links *not* seen.

The immobilization of a protein at a critically low density on an affinity matrix before ‘activation’ with a bifunctional reagent prevents the formation of intermolecular conjugates (Pillai & Bachhawat, 1977, 1978, 1979a; Pillai *et al.*, 1979); for oligomeric proteins it has been demonstrated that this procedure, followed by the examination of the stoichiometry of the cross-linked protomers, can reveal fairly detailed information (Pillai & Bachhawat 1979b; Pillai, 1981) regarding the structure and orientation of subunits. The present study aims to show how the principles of this approach may help predict the detailed three dimensional structure of any protein, *including* single-chain globular proteins.

The task of obtaining homogeneous, determinant-specific antibodies to any crystallizable or non-crystallizable protein has been simplified by the hybridoma approach of Kohler & Milstein (1975). Monoclonal-antibody supports would be essential for such a study: each protein molecule in a single experiment would be immobilized on such a support in an identical fashion (through the single antigenic site being recognized by the monoclonal antibody).

The use of a purified conventional hetero-antiserum would render cross-link analysis impossible. By first immobilizing a protein on its chemically derivatized antibody (attached to a solid matrix), at a suitably low density, it may be made to react with any homo- or hetero-bifunctional, rigid, fixed-chain-length cross-linker under conditions optimized to yield maximum cross-link generation without fear of intermolecular cross-linking taking place.

There are two possible consequences of such a reaction.

(a) Every cross-linker-specific amino acid side chain located at an appropriate distance from a neighbouring relevant side chain will participate in intramolecular cross-links.

(b) All the specific amino acid residues that are not within an appropriate distance from other relevant side chains will react with only one arm of the cross-linker, the other arm being available for reaction with specific labelled quenching reagents. For every residue that is not involved in a 'positive' cross-link, definitive information will thus be available as to its *not* being a particular distance (or less) from all other reactive residues on the surface of the protein. Residues that, although 'exposed', are for reasons of steric hindrance or poor solvation (Glazer, 1976) not reactive, will neither participate in intramolecular cross-links nor be identified by a quench and will thus be automatically excluded from analysis.

Certain problems may arise in the analysis of data thus obtained.

(a) Interactions, both specific and non-specific, between the chemically modified antibody and the protein being studied may sterically 'block' a number of reactive groups on the protein. Since it is likely that any globular protein will have many antigenic sites or epitopes, the routine use of a variety of epitope-specific monoclonal-antibody matrices in parallel would help eliminate matrix effects from analysis.

(b) Competition may occur among cross-linking sites that are equidistant. If a residue A is equidistant from both B and C, depending on their relative reactivities, A and B may be regularly cross-linked and C may be 'spared' or vice versa. If C is so poorly reactive or sterically so hindered that it does not react with cross-linkers and consequently is not identified by a 'quenching' reagent, it will automatically be eliminated from analysis as noted above. If, however, it is freely reactive, hybrid A-B and A-C cross-links may be obtained that will prove directly interpretable. If C is not cross-linked but is identified by a quenching reagent even though it is the same distance from A as B, some errors may enter the analysis. This problem, however, is likely to be rare, but must be viewed as a possible source of error in the overall analysis.

(c) A possible limitation of this approach is that chemical modification itself may cause conformational changes in the protein, and any distance information derived may really be applicable to this particular altered conformational state. However, even in crystallographic approaches, information is derived about a particular conformational state; in general, this limitation is not likely to seriously compromise the overall analysis.

(d) The existence of more than one residue of a modified amino acid in a single peptide fragment could constitute a source of ambiguity; however, since this approach would be utilized only for proteins whose primary structure is known, partial sequencing of 'problem' fragments along with a study of the existing sequence information might help resolve such ambiguities should they arise.

(e) It is quite likely that certain specific chemical modifications may diminish the antigen-binding ability of the antibody being used, if not completely destroy it. An additional advantage of using monoclonal antibodies (apart from the fundamental reason discussed above), would be that in a single fairly simple experiment a number of monoclonal antibodies, all specific for a single protein, could be screened for the effect of a single modification procedure on binding activity. Those antibodies which significantly lose activity after modification may be rejected and the remaining clones retained. A partial loss of activity in any single monoclonal antibody would not constitute a significant problem.

A convenient method of expressing three-dimensional structural information in two dimensional form is the distance matrix, which was initially used for the representation of known structures (Phillips, 1970; Rossman & Liljas, 1974; Dunn & Klotz, 1975) and has more recently been experimented with in predictive approaches (Havel *et al.*, 1972; Kuntz *et al.*, 1979). This representation is likely to be ideal for mapping positive as well as 'negative' distances between amino acid residues.

Eleven of the twenty amino acids commonly found in proteins have side chains for which specific mono- or bi-functional reagents are already available (Glazer, 1976; Uy & Wold, 1977). For these eleven amino acids, at least, there is virtually no limit to the variety of rigid, fixed-length, homo- or hetero-bifunctional reagents that may be synthesized. Certain features that aid cross-link point identification such as cleavability and the presence of reporter groups have already been incorporated in some available reagents (Schramm, 1975; Lutter *et al.*, 1974). A necessary feature would be the ability of the reagent to be used in a two-step (quenching) procedure on the matrix. Easily solvolyzed reagents such as the bis-imidates are likely to prove unsuitable.

It is likely that, for any protein, approx. 30–40%

of its constituent amino acids (Anfinsen & Scheraga 1975; Lee & Richards, 1971) will be exposed and capable of reacting with bifunctional reagents. For every such exposed and reactive residue it should be possible to say either that residue n_i is $x \pm y$ Å from residue n_j (y takes into account the known flexibility of amino acid side chains) or that residue n_i is *not* at a distance $\leq x$ Å from residue n_2, n_3, n_4 etc., as the case may be. If the total number of amino acid residues in a protein is N , a conservative estimate of approx. $N/4 \times N/4$ definitive positive or 'negative' plottable distances should prove available. This magnitude of long-range information is likely to be more than sufficient to resolve the 'multiple minimum' problem. Even a small amount of definite distance information (which may be obtained without ambiguity by this method of low-density immobilization and cross-linking) is likely to prove sufficient to support tertiary-structure predictions.

The information regarding inter-residue distances that will be usually available from this approach will be scalar in nature. With improvements in knowledge of the range of flexibility of amino acid side chains (by spectroscopic techniques, for instance), and of the bond angles generated between cross-linkers and side chains, vectorial information in the form of polar coordinates should also, eventually, prove available.

With techniques now available for obtaining a number of epitope-specific monoclonal antibodies towards any globular protein, such a cross-linking approach, even with existing reagents, is likely to be capable of providing sufficient distance information to help predict the three-dimensional structure of any globular protein. The synthesis of newer, specific, homo- and hetero-bifunctional defined-chain-length cross-linking reagents would increase the amount of distance information thus obtainable.

A large number of biologically interesting membrane-bound proteins, (which include, among others, receptors for hormones, neurotransmitters and antigens), are poorly soluble and particularly difficult to crystallize. Refinements in protein and nucleic acid sequencing methods have simplified the task of obtaining the primary structure of a protein (from the purified protein itself and occasionally from its cloned structural gene). The number of proteins whose three-dimensional structure is known represents but a small fraction of the number of proteins whose primary structure has already been determined.

Most detergents used to solubilize membrane-bound proteins do not compromise immunochemical reactivity. A poorly soluble protein may be conceivably immobilized on an immunoadsorbent matrix in the presence of detergent; the latter could be conveniently washed out before cross-linking and

elution. This approach in combination with predictive methods may prove of practical value in studying a large number of interesting membrane-bound proteins not amenable to study the X-ray-diffraction analysis.

Although the analysis of a protein by such an approach is likely to be a time-consuming and difficult endeavour, the absence of any alternative route towards elucidating the structure of proteins that have not yet proved crystallizable, may make such an effort worthwhile.

I thank Professor B. K. Bachhawat for his encouragement and Mr. E. G. P. Nair for secretarial assistance.

References

- Anfinsen, C. B. & Scheraga, H. A. (1975) *Adv. Protein. Chem.* **29**, 205–300
- Bickle, T. A., Hershey, J. W. B. & Traut, R. R. (1972) *Proc. Natl. Acad. Sci. U.S.A.* **69**, 1327–1332
- D'Albis, A. & Gratzer, W. (1976) *J. Biol. Chem.* **251**, 2825–2830
- Davies, G. E. & Stark, G. R. (1970) *Proc. Natl. Acad. Sci. U.S.A.* **66**, 651–656
- Dunn, J. B. & Klotz, I. (1975) *Arch. Biochem. Biophys.* **167**, 615–618
- Glazer, A. N. (1976) in *The Proteins* (Neurath, H. & Hill, R. H., eds.), vol. 2, pp. 2–103, Academic Press, New York, San Francisco and London
- Guzzo, A. V. (1965) *Biophys. J.* **5**, 809–822
- Hartmann, F. C. & Wold, F. (1967) *Biochemistry* **6**, 2439–2443
- Havel, T. F., Crippen, G. M. & Kuntz, I. D. (1972) *Biopolymers* **18**, 73–81
- Kohler, G. & Milstein, C. (1975) *Nature (London)* **256**, 495–497
- Kuntz, I. D., Crippen, G. W. & Kollman, P. A. (1979) *Biopolymers* **18**, 939–957
- Lee, B. & Richards, F. M. (1971) *J. Mol. Biol.* **55**, 370–400
- Lutter, L. C., Ortanderl, F. & Fasold, H. (1974) *FEBS Lett.* **48**, 288–292
- Matthews, B. W. (1975) *Biochim. Biophys. Acta* **405**, 442–451
- Peters, K. & Richards, F. M. (1977) *Annu. Rev. Biochem.* **46**, 523–551
- Phillips, D. C. (1970) in *British Biochemistry, Past and Present* (Goodwin, T. W. F., ed.), pp. 11–28, Academic Press, London
- Pillai, S. & Bachhawat, B. K. (1977) *Biochem. Biophys. Res. Commun.* **75**, 204–245
- Pillai, S. & Bachhawat, B. K. (1978) *FEBS Lett.* **90**, 51–54
- Pillai, S. & Bachhawat, B. K. (1979a) *FEBS Lett.* **101**, 301–304
- Pillai, S. & Bachhawat, B. K. (1979b) *J. Mol. Biol.* **131**, 877–881

- Pillai, S., Surolia, A. & Bachhawat, B. K. (1979) *Biochim. Biophys. Acta* **580**, 216–221
- Pillai, S. (1981) *Biochem. J.* **193**, 825–828
- Rossmann, M. G. & Liljas, A. (1974) *J. Mol. Biol.* **85**, 177–181
- Schramm, H. J. (1975) *Hoppe-Seyler's Z. Physiol. Chem.* **356**, 1375–1379
- Schulz, G. E., Barry, C. D., Friedman, J., Chou, P. Y., Fasman, C. D., Finkelstein, A. K., Lim, V. I., Ptitsyn, O. B., Kabat, E. A., Wu, T. T., Levitt, M., Robson, B. & Nagano, K. (1974) *Nature (London)* **250**, 140–142
- Steck, T. L. (1972) *J. Mol. Biol.* **66**, 295–305
- Swenson, M. K., Burgess, A. W. & Scheraga, H. A. (1978) in *Frontiers in Physico-chemical Biology* (Pullman, B., ed.), pp. 115–142, Academic Press, New York, San Francisco and London
- Thomas, J. O. & Kornberg, R. D. (1975) *Proc. Natl. Acad. Sci. U.S.A.* **72**, 2626–2630
- Tinberg, H., Lee, C. & Packer, L. (1975) *J. Supramol. Struct.* **3**, 275–283
- Uy, R. & Wold, F. (1977) in *Protein Crosslinking, Biochemical and Molecular Aspects* (Friedman, M., ed.), pp. 169–186, Plenum Press, New York and London