Eur Biophys J (1991) 19:283–286

European Biophysics Journal

Modelling protein three-dimensional structure using tritium planigraphy

A. Gedrovich [1], A. Shishkov [1], V. Goldanskii [1], L. Baratova [2], N. Grebenshchikov [2], and A. Efimov [3]

[1] N. N. Semenov Institute of Chemical Physics, Academy of Sciences of the USSR, Moscow, USSR
[2] A. N. Belozersky Laboratory of Molecular Biology and Bioorganic Chemistry, Moscow State University, Moscow 119899, USSR
[3] Institute of Protein Research, Academy of Sciences of the USSR, Pushchino, Moscow Region, USSR

Received May 21, 1990/Accepted in revised form January 22, 1991

Abstract. We propose the use of data on the topography of the label-accessible surface of a protein molecule obtained by the method of tritium planigraphy as a criterion for choosing the optimal intermediate arrangements of *alpha*-helices in globular proteins so as to model their three-dimensional structures. This approach has been used for modelling the three-dimensional structure of parvalbumin III from pike. The proposed model has been compared with high-resolution X-ray structural data for a related protein, paryvalbumin from carp. The possibilities and limitations of this approach are discussed.

Key words: Pike parvalbumin – Tritium planigraphy – Three-dimensional modelling

Introduction

In 1976 we described a method for obtaining tritium-labelled biologically active substances. The method is based on the high reactivity of tritium atoms produced via thermal dissociation of tritium molecules on a tungsten filament heated to 2000 K (Shishkov et al. 1976). The intramolecular distribution of the tritium in the *N*-terminal part of sperm-whale of myoglobin was in good agreement with X-ray structural data for this protein (Goldanskii et al. 1982). This allowed us to work out a new experimental method for studying the topography of the surface of macromolecular structures – the so-called tritium planigraphy method (Goldanskii et al. 1988).

This method has been used by us and other researchers for studying the topography of biological macromolecular and supramolecular systems (Alyonysheva et al. 1987; Jusupov and Spirin 1986; Baratova et al. 1990). In the present work we suggest a further use for thermally activated tritium atoms. We propose that using the tritium planigraphy method to obtain experimental data on the topography of the label-accessible surface of a protein molecule, in conjunction with theoretical approaches, enables one to choose an optimal way for modelling protein three-dimensional structure.

Offprint requests to: A. Shishkov, N. N. Semenov Institute of Chem. Phys., USSR Academy of Sciences, Ul. Kosygina 4, 117334, Moscow, USSR

Materials and methods

Parvalbumin III from pike was kindly provided by Dr. V. N. Medvedkin, Protein Research Institute, Academy of Sciences of the USSR, Pushchino.

The label was introduced as described by Goldanskii et al. (1988). Parvalbumin solutions (2 mg/ml) in a volume of 2 ml in 0.05 *M* ammonium bicarbonate were sprayed on the liquid nitrogen-precooled inner surface of the vessel. Elimination of exchangeable (labile) tritium from the labelled protein samples was performed by gel-filtration on a 45 × 1 cm G-25 (fine) Sephadex column (Pharmacia, Sweden). Elution was performed with 0.05 *M* NH_4HCO_3.

Digestion with protease V8 from *St. aureus* was carried out at an enzyme/substrate ratio of 1/50 at 37 °C in 0.1 *M* *N*-methylmorpholine acetate buffer, pH 8.0, for 2.0 h.

HPLC of hydrolysates and analysis of peptides were performed as described by Goldanskii et al. (1988).

The secondary structure of parvalbumin III from pike was predicted by the statistical method of Chou and Fasman (1978) and by the method proposed by Lim (1974).

Results and discussion

A search for approaches for the prediction of a protein's three-dimensional structure continues to command intense interest. Three major groups of methods are used for predicting protein tertiary structure: *(1)* using modelling based on the known three-dimensional structures of homologous proteins; *(2)* using methods for predict-

ing secondary structures, followed by the assembly of these units into a compact tertiary structure; *(3)* using empirical energy functions ab initio for deriving a tertiary structure of minimum potential energy.

With some exceptions, each of the approaches permit some variants of the spatial arrangement of secondary structure elements and therefore it is necessary to choose among two or more alternative models of packing during each stage of the assembly.

We propose that a comparison of theoretical models (or their parts) obtained by one of the methods with experimental data on the topography of the protein molecule surface could facilitate the choice of optimal modes of packing on all the stages of assembly. Since the method of tritium planigraphy allows us to obtain such information, we ventured to make an attempt at solving this problem. In the general form we suggest the following way of modelling three-dimensional structure: *(i)* to determine the profile of the accessibility of amino acid residues in a protein to tritium-labelling; *(ii)* to predict secondary structure elements using any algorithm; *(iii)* to search such mode(s) of spatial arrangement of protein *alpha*-helices so as to provide the surface location of protein sites most accessible to tritium labelling. It should be noted that elements of the secondary structure are considered as absolutely rigid.

To test this approach we modelled the three-dimensional structure of parvalbumin III from pike. It is a globular protein from the family of Ca^{2+}-binding proteins. It consists of 108 amino acid residues, i.e. it is relatively small (Gerday 1976). There are nine glutamic acid residues in the protein and it was for this reason that we chose protease V8 from *St. aureus* for fragmentation of the native protein. For analysis of the intramolecular distribution of the tritium label in the protein we used fragmentation of tritiated protein followed by amino acid analysis of homogeneous peptides with a simultaneous radioactivity count for each type of amino acid residue. The optimal solution of this problem is possible if the fragmentation of the initial protein into short peptides, with a minimum of amino acid repeats in each peptide, is achieved. If the peptide contains two or more of the same amino acid residue we obtain only the average value of tritium label incorporation for that particular residue.

In our case, individual, i.e. once met, residues in peptides of the V8 protease hydrolysate of pike parvalbumin constituted only ~20% of the sum of all residues present in the analysed peptides. However, such average data on the intramolecular distribution of the label in peptides is sufficient to illustrate our approach in constructing the three-dimensional structure of a protein, since this information is supplemented with the knowledge of Ca^{2+}-binding sites located in the region of *beta*-turns and with the prediction of a high percentage of *alpha*-helices, which allows one to represent the greater part of the protein structure as rigid cylinders. Therefore we have used only one method of protein fragmentation and have worked with large peptides such as V1–V4.

The experimental profile of amino acid accessibility to labelling was determined from the data on the intramolecular distribution of the tritium label in the protein (Fig. 1). One can see that virtually all amino acid residues in the protein are accessible to tritium labelling to some extent and this is quite natural. For instance, the percentage of lysozyme, myoglobin or ribonuclease residues that would be accessible to contact with a water-sized sphere (with radius, $r = 0.14$ nm) calculated by the method of Lee and Richards (1971) is also not equal to zero.

In the present attempt to model three-dimensional protein structure we used an approach based on the prediction of secondary structure units followed by assembly of these units into a compact structure. Using the methods of Lim (1974) and Chou and Fasman (1978), we have predicted, six *alpha*-helical regions: 5–20, 25–33, 42–51, 57–70, 79–88 and 99–108. These *alpha*-helices will be noted as A-, B-, C-, D-, E- and F-helices (Fig. 1). We shall designate the predicted beta-turns between *alpha*-helices as loops AB (21–24), BC (38–41), CD (52–56), DE (71–76) and EF (89–95). The predicted *alpha*-helices include 69 amino acids out of 108, which amounts to 63.9% of the sequence. This value is high compared with an earlier circular dichroism estimate (36.5%) (Gosselin-

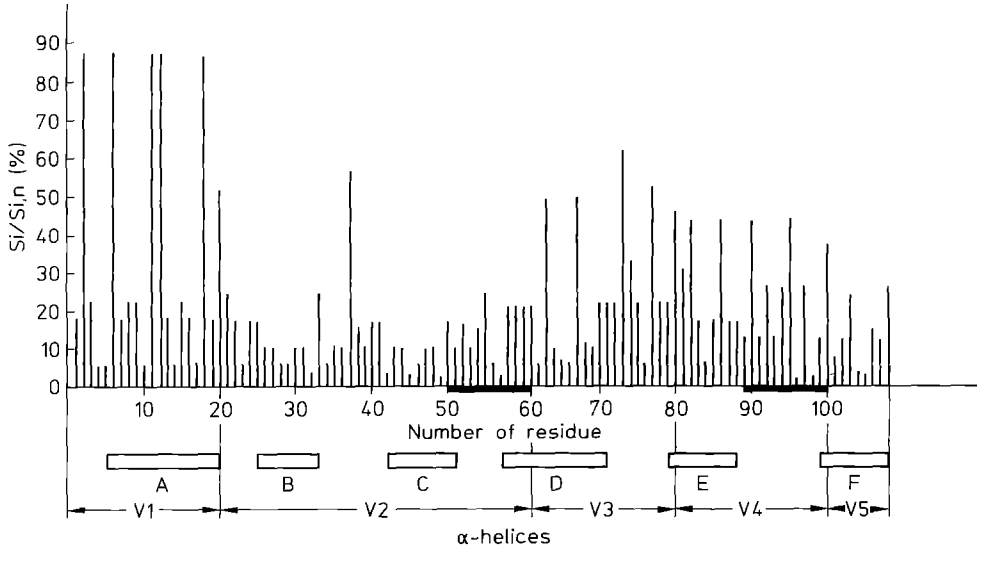

Fig. 1. Intramolecular distribution of the tritium label in pike parvalbumin III. (Si/si, n) % – ratio of the accessible surface area of the amino acid residue (CH-bound) in the protein to that in a *Gly-X-Gly* peptide (the model for 100%-accessibility of the residue (Shrake and Ruplex (1973)). ▭ – *alpha*-helic regions; ← → – V1–V5-protease V8 peptide; ▬ – Ca^{2+}-binding sites

Rey et al. 1973) but is in good agreement with later data of Chang et al. (1978) showing 59% helix for carp parvalbumin. This is closely related to pike parvalbumin (Chothia and Lesk 1986). Some discrepancy between prediced *alpha*-helix content and that measured by CD is not unusual.

For modelling the three-dimensional structure of pike parvalbumin III, we adopted the method proposed by Ptitzyn and Rashin (1975) which involves as its most significant part a search for contacts between the neighbouring *alpha*-helices. Besides, the *alpha*-helices ought to be packed so that amino acid residues of *alpha*-helices with minimal label accessibilities were in the regions of contact of these helices, but not on their surface. *Alpha*-helices were approximated by cylinders with parameters proportional to those of the *alpha*-helix: helix radius 5A, pitch 3.0×1.5 A.

The degree of tritium-accessibility of the amino acid residues of pike parvalbumin *alpha*-helices varies. The mean value for label accessibility of the residues in A-*alpha*-helix is 35.1%, in B-11.0%, in C-8.4%, in D-17.8%, in E-23.4% and in F-15.6%. The values for the B- and C-helices are lower than those in other *alpha*-helices and thus one could expect that they (and, perhaps, partly the D-helix) must be involved in forming the protein molecule's hydrophobic part. Indeed, one of the characteristic features of the C- and D-helices is the clusters of hydrophobic amino acid residues, which have much less of the tritium-label than other amino acids of these helices. On the face of it, in their interface the contact of pairs *Phe* 69-*Val* 42, *Leu* 66-*Phe* 46, *Val* 65-*Val* 45 and *Leu* 62-*Ile* 49 are the most suitable candidates which satisfy the requirements for the minimum accessibilities of these residues to the tritium. However, if this is the case then the positions of *Asp* 50, *Glu* 58 and *Glu* 61, which participate in coordination of the Ca^{2+}-ion on the opposite sides of an octahedron, symbolized by the coordinate sphere of the Ca^{2+}-ion (Tufty and Kretsinger 1975) must also be correct. We find that *Asp* 50 cannot co-ordinate Ca^{2+}, because the D-helix will screen it from other residues involved in binding Ca^{2+}. Therefore we suggest that the C- and D-helices form their contact in the following way: *Val* 45-*Ser* 68, *Ile* 49-*Phe* 64 (Fig. 2A).

Such mode of packing is in good agreement with tritium accessibility data on the polypeptide chain: a hydrophobic cluster, consisting of *Phe* 46, *Leu* 66, *Val* 65 and *Leu* 62 is formed on the surface of the C- and D-helices. To some extent residues *Val* 42 and *Phe* 69 also take part in forming this cluster.

The E-helix can be placed on the surface of the double complex of C- and D-helices so that the contacts *Ala* 88-*Leu* 62, *Ala* 87-*Val* 65, *Phe* 84-*Leu* 66 and, which is less likely, *Ala* 83-*Val* 42 are formed (Fig. 2B).

Now the F-helix can be placed on the surface of the triple complex of C-, D- and E-helices, and there is one possible way of doing it. As a result the contacts *Val* 42-*Val* 105 and *Phe* 46-*Phe* 101 are formed (Fig 2C). There is a second Ca^{2+}-binding site on the protein in the EF-loop region (residues 89 to 100). And finally, it is necessary to find the contact between the A- and B-helices, and to bind them to the CDEF-helix complex. Considering

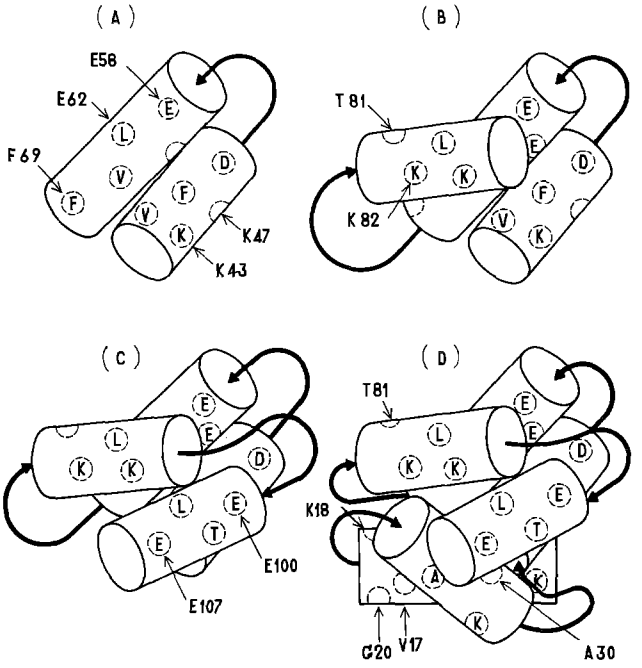

Fig. 2A–D. The "step-by-step" packing of *alpha*-helices and a schematic model of the three-dimensional structure of pike parvalbumin III

the distribution of the label in the A- and B-helices, we propose that the contact between these *alpha*-helices is established through interaction of residues *Ile* 10-*Val* 32, *Ala* 13-*Phe* 28 and *Leu* 14-*Phe* 29. Although the double complex of A- and B-helices can, in principle, contact the CDEF-four-helix complex via interaction with the CD-helices or the EF-helices, we suggest that the former takes place (Fig. 2D). This is because of the distribution of the label in the F-helix (Fig. 1), residues *Phe* 101, *Leu* 104 and *Val* 105 are least accessible to tritium labelling. They are grouped in a single region of the helix surface and form a contact surface of the F-helix in the CDEF-four-helix complex (Fig. 2D). Amino acid residues most accessible to tritium atoms are *Ala* 108, *His* 106, *Thr* 103 and *Glu* 100. They are located on the other side of the helix, which is "outward" with respect to the other parts of the globule.

This type of accessibility is typical for the amino acid residues in the helix exposed to the solvent. So it is logical to bind the AB-helices to the CDEF-four-helical complex via contact of the N- and C-termini of C- and D-helices with the B-helix. In this case the following contacts are possible: *Ala* 30-*Ala* 108, *Ala* 70-*Leu* 14 and *Phe* 29-*Val* 42. We have predicted five beta-turns in pike parvalbumin III designated as the AB-, BC-, CD-, DE- and EF-loops. The best label accessibilities are in the residues of the DE- and EF-loops (28.3 and 25.8%, respectively). The CD-loop is the least accessible to labelling (14.6%). As noted above, two Ca-binding sites are located in the region of the CD- and EF-loops (residues 50–61 and 89–100, respectively). The constants of Ca^{2+}-binding for the pike parvalbumin III also confirm the presence of two Ca^{2+}-binding regions but show no difference in their affinity for Ca^{2+}-ions (Permyakov et al. 1983). The three-dimensional model obtained is in good agreement with

the well-known three-dimensional structure for the related protein, parvalbumin from carp (Kretsinger and Nockolds 1973; Moews and Kretsinger 1975). For comparison we have constructed a pike parvalbumin III model on the basis of high-resolution X-ray structural data for the related (62% homology) protein, carp parvalbumin. For assembly we used the contacts between *alpha*-helices given in Moews and Kretsinger (1975). A comparison of both models showed similarity in the three-dimensional organization of molecules. Yet in the absence of high resolution X-ray structured data for the pike parvalbumin III, we can not draw a final conclusion about the prospects of our approach for modelling protein three-dimensional structures with the aid of thermally activated tritium atoms. The present work is an initial step in this direction. Manual exhaustion of variants to pack up elements of the secondary structure might be adopted only for low molecular weight proteins. Appropriate computer programs must certainly be developed for the purpose.

In conclusion, some general considerations on the traditional theoretical approaches for modelling biopolymer three dimensional structures may be made. All these approaches are predicated on the assumption that a polypeptide chain has *already been synthesized*; the focus is on the interactions of chain fragments with one another and with the environment (water). The temporal factor is omitted from consideration altogether. Yet in real conditions, polypeptide chain synthesis on the ribosome is a gradual process; during translation, the growing chain is transferred into the "medium" and, in all likelihood, structural domains may already arise at this stage.

Acknowledgements. The authors would like to thank Dr. V. Medvedkin for supplying the pike parvalbumin III, Dr. E. Permyakov for his critical remarks about the manuscript, and I. S. Kochubey for his help in editing the English text.

References

Alyonysheva TN, Kuryatov AB, Antropova LP, Shemyakin VV, Neiman LA, Tsetlin VI, Ivanov VT (1987) Tritium planigraphy study on the bacteriorhodopsin topography. Bioorganitcheskaya Khimia 13:898–907 (in Russian)

Baratova LA, Grebenshchikov NI, Shishkov AV, Kashirin IA, Radavsky YS, Jarvekulg LK, Saarma MJ (1990) The topography of the surface of the potato virus X: tritium planigraphy and immunological analysis. J Gen Virol (in press)

Chang CT, Wu C-SC, Yang JT (1978) Circular dichroic analysis of protein conformation: inclusion of the β-turns. Anal Biochem 91:13–31

Chothia C, Lesk AM (1986) The relation between the divergence of sequence and structure in proteins. The EMBO J 5:823–826

Chou PV, Fasman GD (1978) Empirical prediction of protein conformation. Ann Rev Biochem 47:251–276

Gerday C (1976) The primary structure of the parvalbumin III of pike. Eur J Biochem 70:305–318

Goldanskii VI, Rumyantsev YuM, Shishkov A, Baratova LA, Belyanova LP (1982) Study of the three-dimensional structure of proteins by means of tritium labelling. II. Intramolecular distribution of tritium in the N-terminal part of myoglobin and the tertiary structure of protein. Molekularnaya Biologiya 16:528–534 (in Russian)

Goldanskii VI, Kashirin IA, Shishkov AV, Baratova LA, Grebenshchikov NI (1988) The use of thermally activated tritium atoms for structural-biological investigations: topography of the TMV protein-accessible surface of the virus. J Mol Biol 201:567–574

Gosselin-Rey C, Bernhard N, Gerday C (1973) Conformation and immunochemistry of parvalbumin III from pike white muscle: modification of the arginine with 1,2-cyclohexanedione. Biochim Biophys Acta 303:90–114

Jusupov MM, Spirin AS (1986) Are the proteins between ribosomal subunits? Hot tritium bomardment experiments. FEBS Lett 197:229–233

Kretsinger RH, Nockolds CE (1973) Carp muscle calcium binding proteins. J Biol Chem 248:3313–3326

Lee B, Richards FM (1971) The interpretation of protein structures: Estimation of static accessibility. J Mol Biol 55:379–400

Lim VI (1974) Algorithms for prediction of *alpha*-helical and β-structural regions in globular proteins. J Mol Biol 88:872–894

Moews PC, Kretsinger RH (1975) Refinement of the structure of carp muscle calcium-binding parvalbumin model building and difference fourier analysis. J Mol Biol 91:201–228

Permyakov EA, Medvedkin VN, Kalinichenko CP, Burstein EA (1983) Comparative study of physico-chemical properties of two pike parvalbumins by means of their intrinsic tyrosine and phenylalanine fluorescence. Arch Biochem Biophys 227:9–20

Ftizyn OB, Rashin AA (1975) A model of myoglobin self-organization. Biophys Chem 3:1–20

Shishkov AV, Filatov ES, Simonov EF, Unukovich MS, Goldanskii VI, Nesmeyanov AN (1976) Obtaining tritium labeled biological active compounds. Dokl Akad Nauk USSR 228:1237–1239

Shrake JA, Rupley JA (1973) Environment and exposure to solvent of protein atoms. Lysozyme and insulin. J Mol Biol 79:351–371

Tufty RH, Kretsinger RH (1975) Troponin and parvalbumin calcium binding regions predicted in myosin light chain and T4 lysozyme. Science 187:167–169