Biophysical Chemistry 29 (1988) 219–224
Elsevier

BPC 01206

Minireview

Methods of molecular modelling of protein-protein interactions

Piotr Zielenkiewicz and Andrzej Rabczenko

Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Rakowiecka 36, 02532 Warsaw, Poland

Received 9 March 1987
Accepted 26 October 1987

Protein-protein interaction; Molecular modeling; Quaternary structure

This article reviews briefly theoretical methods attempting to predict the structure of protein aggregates from the structural features of their subunits. The authors discuss the problems of the solvent effect and the formation of protein structure. The existing methods of quaternary structure prediction are presented and an attempt at their classification is made at the end of this review.

1. Scope

Protein-protein recognition processes, leading to the formation of noncovalently bound protein aggregates (protein quaternary structure), are one of the major factors involved in the regulation of metabolic pathways in living cells as a result of changes in the functional properties of proteins with association state. Several excellent reviews on various aspects of the structure and functioning of multisubunit proteins have already been published [1–6].

The aim of this article is to review briefly theoretical methods attempting to predict the structure of protein aggregates from the structural features of their subunits.

Although the physical principles of protein-protein recognition are quite well understood their application in predictive methods is still not straightforward because of the complexity of the systems involved. The need to examine and to compare a large number of different mutual orientations of subunits in order to determine the most favorable mode of interaction requires the use of simplifying assumptions for making the computations for one specific orientation very fast. The set of simplifying assumptions employed is one of the most distinctive features of the predictive methods described below and, in our opinion, the most effective procedure for their classification. In the light of this point of view, in the following sections we describe firstly assertions underlying the molecular modelling of proteins in general, assumptions common to the methods of modelling protein-protein interactions and finally list the methods followed by an attempt at their classification according to the assumptions concerned.

Correspondence address: P. Zielenkiewicz, Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Rakowiecka 36, 02532 Warsaw, Poland.

2. Preamble to introduction

The factors dictating the structures of proteins are strongly influenced by environmental conditions. The so-called 'solvent effect' is in fact the principal driving force in protein folding and association. Almost all the properties of a protein, exclusive of the amino acid sequence, vary with changes in the environment. In the modelling of a protein structure, which we shall deal with herein, the influence of solvent cannot be established quantitatively as yet. The evaluation of this effect is always approximate and often based on the

knowledge of the experimentally recognised native structures of proteins. Thus, it is important to bear in mind that the considerations presented in this paper are applicable to proteins under conditions close to physiological and that the problem posed by the molecular environment, although not always clearly stated, is always present.

3. Introduction

The term protein-protein interaction is used to describe the formation of noncovalently bound protein subunit aggregates. This reflects the point of view of biochemists who observe the formation of protein complexes with a defined stoichiometry when mixing subunits. However, an investigator dealing with the modelling of protein structure can take a totally different stance. Folding of the polypeptide chain into its native conformation can also be treated as a protein-protein self-interaction. Formation of the tertiary and quaternary structure of a protein is governed by the same rules – the difference in modelling these two processes being technical rather than conceptual.

The sequence describing the hierarchy of protein anatomy, i.e., primary structure, secondary structure, supersecondary structure, protein domain, tertiary structure, quaternary structure, does not correspond to the order of events in the formation of the native structure. Many attempts have been made to predict protein secondary structure from the local amino acid sequence. The limited success [7] of these methods results from the fact that the tendency of a polypeptide chain segment to exist in a given secondary structure depends strongly upon its environment. Recent studies [8,9] show that the formation of secondary structures is the result of interactions within the entire protein domain, in other words, the same local sequence can code different secondary structures. It follows that the structures of entire protein domains should be predicted by way of a one-step procedure with the resulting secondary and supersecondary structures being treated only as useful for describing the anatomy of the domain (for a review, see ref. 10). On the other hand, the two final steps in the formation of a protein structure, i.e., domain association into a tertiary structure in multidomain proteins and subunit association into a protein complex, seem to be very similar. It has been shown that, after proteolytic cleavage between the domains, they can fold up independently [11] and aggregate into the native tertiary structure acting as subunits in a protein complex. Multidomain proteins and protein complexes share in common both functional (e.g., flexibility) and structural (e.g., complementary interfaces) features.

To sum up, the modelling of a protein structure can be divided into two steps:

(i) Prediction of domain structure on the basis of the amino acid sequence of the entire polypeptide chain;

(ii) Stereospecific aggregation of domains or subunits.

In this paper we shall only deal with the second step. Although it is generally recognised that the functional characteristics of native proteins depend strongly on their quaternary structure (for a review, see refs. 2 and 6), there exist only a few theoretical methods for the description and prediction of the modes of association of proteins. In this paper, we shall review some useful methods for the prediction of protein quaternary structure and mapping of the intermolecular contacts at interfaces in multisubunit proteins. These methods can also be applied for the description of domain-domain interactions.

4. Assumptions

Proper computer simulation of protein association should take into account the interactions between all individual atoms of the subunits, their interactions with the environment and changes in both subunits and environment on aggregation. A procedure including these factors should be repeated for many different mutual orientations of the subunits in order to establish which is the most favorable energetically. At the present stage, this is an unresolved problem owing to the complexity of the calculations involved.

In order to simplify the calculations in modelling subunit interactions three major assumptions are made:

(i) The tertiary structure of the subunits is known. X-ray crystallography is the only basis for what we know about the tertiary structure of proteins. Using crystallographic coordinates as input comprises the hidden assumption that the structure in solution is the same as that in the crystalline state (for a discussion of the validity of this assumption, see ref. 12).

(ii) Association does not perturb the tertiary structures of subunits, or at least conformational changes of molecules are sufficiently small not to affect the process of mutual recognition. This assumption appears to be justified by the experimental evidence available [13–16].

(iii) The surface properties of subunits are the most important factor in their interaction. Thermodynamic considerations and analysis of the known structures of protein complexes [17,18] have shown that in order for two subunits to associate they must have surface areas complementary to each other. Geometrical complementarity of subunit surfaces is accompanied by proper positioning of polar and nonpolar groups so as to form all possible hydrogen bonds, salt bridges and hydrophobic contacts (which can be termed complementarity of interactions). The concept of surface complementarity renders the problem of predicting quatenary structure manageable by allowing one to resolve extensive computation of the interaction energies into comparison of the geometrical features of two surfaces. In fact, most of the methods make use of the 'complementarity rule' as one of the basic assumptions.

Some other assumptions, often applicable only to specific systems, are used as well. Below, we describe methods of modelling protein associations, followed by an attempt at their classification according to the assumption used.

5. Methods

A simple amino acid substitution (Glu → Val) in position 6 of the β-hemoglobin chain causes a serious 'molecular disease' – sickle cell anemia [19]. One of the most important aspects of this disease is that deoxygenated sickle cell hemoglobin aggregates to form helical fibers. The basic geometry of these fibers is known from electron microscopy [20]. Levinthal et al. [21], using computerised model building, proposed an arrangement for the molecules in the fibers that satisfies all their reported properties. The authors assumed that at least one β_6 Val is either in or very close to a contact region. This assumption, together with taking into account the known geometry of fibers, led to a decrease in the number of orientations for which the best packing was searched. Levinthal et al. were the first to use polar coordinates as the most convenient system for defining the relative orientation of two protein molecules.

Greer and Bush [22] introduced a quantitative function (equivalent to the molecular surface proposed by Richards [23]) which describes the accessibility of solvent to a macromolecule's surface. This function was used to demarcate complementary sites in the contacts between subunits and for visualisation of both subunit contacts and features of subunit surfaces. This method was applied to a known intersubunit contact, $\alpha_1\beta_1$ in methemoglobin, and to a hypothetical contact in the β-hemoglobin tetramer. It was assumed that the most stable dimer β_2 is similar in structure to $\alpha_1\beta_1$. Such a structure of β_2 was generated by rotating the β-chain by a least-squares fit of homologous α-carbons onto the α_1-chain. By using the proposed function, a complementary fit of subunit surfaces in the hypothetical structure of β_2 was demonstrated.

Salemme [24] suggested a hypothetical structural complex of cytochromes c and b_5. It was found that, in the region where the heme is most accessible to solvent, there exists in both cytochromes a group of charge-conserved residues. Least-squares planes through the charge-invariant groups were calculated and then the positions of these groups relative to the planes were plotted. This procedure provided topological surface charge maps in comparable reference frames and allowed examination of the quality of the complementary fit. In the next step checking of the structural complementarity in the resultant complexes was performed by using rigid space-filling models. The assumption that highly conserved charged groups are important in interactions of cytochrome c with other proteins provided the basis for two

subsequent predictions of complex structure: cytochrome *c*-cytochrome *c* peroxidase [25] and cytochrome *c*-flavodoxin [26]. In these studies, models of complexes were constructed with the aid of a computer graphics display system. The hypothetical structure of the cytochrome *c*-flavodoxin complex was used later to assess the role of complementary electrostatics in the preorientation of molecules and enhancing reaction kinetics [27].

Computer graphics was also used to predict the dimer structure of βB_p-crystallin. Wistow et al. [28], on the basis of the sequence homology between β- and γ-crystallins, predicted the tertiary structure of βB_p-crystallin. In solution it forms a stable dimer. Using interactive computer graphics the authors proposed two possible dimer structures.

Wodak and Janin [29] proposed an automatic procedure for generating possible modes of protein-protein association. These authors used a simplified version of the model of Levitt [30] in which each residue is replaced by one effective interaction center. This method was applied to the study of the trypsin inhibitor-trypsin complex as a model system. Possible modes of interaction were generated systematically. For each mode, the areas buried at interfaces and nonbonding interaction energies were evaluated approximately; more detailed calculations were performed only for selected orientations. This has led to nine structures with parameters similar to those of the native complex (including the native complex structure).

A similar method was suggested by Rashin and Yudman [31]. Following the assumptions that the self-assembly of proteins may be divided into two steps, viz., (i) rough recognition in which one or a few sites are outlined; (ii) fitting of these possible sites, and that the rough recognition is provided by the surface distribution of hydrophobic groups, the energy of dehydration was calculated as the first step. Again, the model of Levitt was used to simplify calculations. In the second step the sum of hydrophobic, van der Waals and hydrogen bond interaction energies was calculated.

Another method which allows one to search over many different subunit orientations was suggested by Zielenkiewicz and Rabczenko [32,33]. This procedure is based on projections of the surface atoms. Certain features (e.g., hydrophobicity and hydrophilicity) were assigned to nonhydrogen atoms to indicate the specificity of interactions. Surface atoms were projected onto a plane scaled by a network of squares. That part of the network comprising the entire area of projection was then described by two matrices representing the three-dimensional surface topology and characteristic features of the atoms. A measure of the complementarity of two surface representations including both topological and interaction complementarity was defined. The results obtained for the well-known autoassociation of insulin demonstrated the applicability of this method to the prediction of possible interaction modes between macromolecules. The computational procedure is efficient enough not to require any additional assumptions about the system taken into account, although diversification of the types of interaction makes insertion of specific information possible. This method was used to predict the structure of βB_p-crystallin [34]. The results obtained using this prediction method are different from those described by Wistow et al. [28].

Recently, a method of molecular cartography [35] was introduced in order to quantitate the topographic structure of a protein surface. This was used to investigate the local and global topography of reported antigenic regions on the surface of myoglobin and lysozyme. Although extension of this method to make it applicable to the study of other types of protein-protein interactions is not straightforward, it does provide a new, interesting way of representing the molecular surface. Inclusion of the hydrophobicity and electrostatic potential of local areas on a protein surface which is intended by the authors, can make molecular cartography a useful tool in the examination of protein-protein recognition as well.

A new efficient algorithm for determining surface shape complementarity was recently proposed by Conolly [36]. Several useful concepts for measuring the shape of protein surface regions are introduced. The idea, to predict the structure of protein complexes, is to match complementary patterns of knobs and depressions on subunit surfaces. The predictive method is sufficiently

general not to employ any simplifications about the particular protein-protein system or any other approximations except the widely accepted assertions listed in section 4. The method uniquely predicted the $\alpha_1\beta_1$ interface in the hemoglobin tetramer, but failed to predict the structure of the trypsin inhibitor-trypsin complex.

Many of the methods outlined above may also serve as a tool for analysing subunit contacts in known complex structures. Some other useful procedures have been proposed for visualising the orientation of molecules and subunit interfaces on paper or a computer screen [37–39]. These methods enable more structural information to be conveyed than from a list of residues in contact (which is, apart from drawings, a standard form of presentation).

6. Concluding remarks

In general, the above-mentioned methods may be grouped into two major categories according to the type of simplifying assumptions employed.

The first group consists of those in which only a limited number of mutual subunit orientations is considered. These methods allow for detailed calculations of intersubunit interactions and may take into account individual contacts at subunit interfaces. In the simplest case, when the complex structure is known from X-ray crystallography, such methods serve as a tool for the description of structural features of a complex [37–39]. Methods also belonging to this group include those in which we take into account additional information (other than tertiary subunit structure) about the system which leads to the required reduction in the number of orientations of molecules considered. In some studies [21,24] the calculations made use of known properties of the system, however this additional information may also be hypothetical [22]. Computer graphics should be included in this group of methods as well. Although no additional assumptions are necessary here, most mutual orientations cancel out on the basis of subjective visual inspection.

The second group comprises those in which different mutual orientations of subunits are searched systematically in order to determine the most favorable mode of association. Scanning of a large number of configurations becomes possible through simplification of the calculations of intersubunit interactions; these methods allow one to select a small number of interfaces for which more extensive and detailed calculations should be made (a method from the first group may be chosen for this purpose). Methods falling within this category include those of Wodak and Janin [29], Rashin and Yudman [31] and Zielenkiewicz and Rabczenko [32–34].

To summarize, at the present stage, computer modelling of protein associations is possible only if crucial simplifying assumptions are made. If we look at procedures allowing the systematic evaluation of possible modes of interaction, these being the most interesting in their applicability to different systems, we are in fact in the position of searching for such a level of approximation that will not yet lead to oversimplification. Nevertheless, the optimistic issue is that methods for the prediction of quaternary structure provide correct descriptions of the model systems and reasonable hypothetical structures. This is far beyond the progress made in predictions of protein tertiary structure.

References

1 F.K. Friedman and S. Beychock, Annu. Rev. Biochem. 48 (1979) 217.
2 C. Frieden, Annu. Rev. Biochem. 40 (1971) 653.
3 B.W. Matthews and S.A. Berhnard, Annu. Rev. Biophys. Bioeng. 2 (1973) 257.
4 R. Jaenicke, Angew. Chem. 23 (1984) 395.
5 R. Jaenicke, Biophys. Struct. Mech. 8 (1982) 231.
6 I.M. Klotz, D.W. Darnall and N.R. Langerman, in: The proteins, 3rd edn., vol. 1, eds. M. Neurath and R.L. Hill (Academic Press, New York, 1975) p. 293.
7 K. Nishikawa, Biochim. Biophys. Acta 748 (1983) 285.
8 W. Kabsch and C. Sander, Proc. Natl. Acad. Sci. U.S.A. 81 (1984) 1075.
9 D. Plochocka, J. Kosinski and A. Rabczenko, Acta Biochim. Pol. 33 (1986) 109.
10 J.S. Richardson, Adv. Protein Chem. 34 (1981) 167.
11 C. Ghelis and J. Yon, Protein folding (Academic Press, New York, 1982).
12 B.W. Matthews, in: The proteins, 3rd edn., vol. 3, eds. M. Neurath and R.L. Hill (Academic Press, New York, 1975) p. 403.

13 J. Baldwin and C. Chotia, J. Mol. Biol. 129 (1979) 175.
14 W.S. Bennet and T.A. Steitz, J. Mol. Biol. 140 (1980) 211.
15 L.S. Hsu and K.E. Neet, J. Mol. Biol. 97 (1975) 351.
16 R. Huber, D. Kukla, W. Bode, P. Schwager, K. Bartels, J. Deisenhofer and W. Steigeman, J. Mol. Biol. 89 (1974) 73.
17 C. Chotia and J. Janin, Nature 256 (1975) 705.
18 P.D. Ross and S. Subramian, Biochemistry 20 (1981) 3096.
19 L. Pauling, H.A. Itano, S.J. Singer and I.C. Wells, Science 110 (1949) 543.
20 B. Magdoff-Fairchild and P. Swerdlow, Nature 239 (1972) 217.
21 C. Levinthal, S.J. Wodak, P. Kahn and A.K. Dadivanian, Proc. Natl. Acad. Sci. U.S.A. 72 (1975) 1330.
22 J. Greer and B.L. Bush, Proc. Natl. Acad. Sci. U.S.A. 75 (1978) 303.
23 F.M. Richards, Annu. Rev. Biophys. Bioeng. 6 (1977) 151.
24 F.R. Salemme, J. Mol. Biol. 102 (1976) 563.
25 T.L. Poulos and J. Kraut, J. Biol. Chem. 255 (1980) 10322.
26 R.P. Simondsen, P.C. Weber, F.R. Salemme and G. Tollin, Biochemistry 21 (1982) 6366.
27 J.B. Matthews, P.C. Weber, F.R. Salemme and F.M. Richards Nature 301 (1983) 169.
28 G. Wistow, C. Slingsby, T. Blundell, H. Driessen, W. DeJong and H. Bloemendal, FEBS Lett. 133 (1981) 9.
29 S. Wodak and J. Janin, J. Mol. Biol. 124 (1978) 323.
30 M. Levitt, J. Mol. Biol. 104 (1976) 59.
31 A.A. Rashin and B.M. Yudman, FEBS Lett. 101 (1979) 6.
32 P. Zielenkiewicz and A. Rabczenko, J. Theor. Biol. 111 (1984) 17.
33 P. Zielenkiewicz and A. Rabczenko, J. Theor. Biol. 116 (1985) 607.
34 P. Zielenkiewicz and A. Rabczenko, Acta Biochim. Pol. 34 (1987) 87.
35 D.W. Fanning, J.A. Smith and G.D. Rose, Biopolymers 25 (1986) 863.
36 M.J. Conolly, Biopolymers 25 (1986) 1229.
37 R.J. Feldman, D.H. Bing, B.C. Furie and B. Furie, Proc. Natl. Acad. Sci. U.S.A. 75 (1978) 5409.
38 P.K. Weiner, R., Langridge, J.M. Blaney, R. Schaefer and P.A. Kollman, Proc. Natl. Acad. Sci. U.S.A. 79 (1982) 3754.
39 E.J. Milner-White, Biochem. J. 205 (1982) 353.