TOWARDS THEORETICAL PREDICTION OF PROTEIN QUATERNARY STRUCTURE

A. A. RASHIN and Bella H. YUDMAN
Institute of Biological Physics, USSR Academy of Sciences, 142292, Poustchino, Moscow Region, USSR

Received 5 March 1979

1. Introduction

Many protein complexes are formed by the self-assembly from preformed subunits. This fact is the basis of our present attempt to calculate protein quaternary structure from the known tertiary structure of its subunits. It assumes that globular parts of proteins participating in protein–protein interactions do not change significantly upon the assembly. Such a belief is justified by the experimental evidence [1–4] and thus is a reasonable working hypothesis.

We have assumed that protein self-assembly should include at least two main steps to be effective:

1. Rough recognition at which one or a few possible interaction sites are outlined by interactions not demanding any detailed fitting of subunit surfaces;
2. Fitting of these possible sites which leads to the most stable structure.

The stability of many known protein complexes may depend upon hydrophobic interactions [5–11]. Thus we have also assumed that the rough recognition may be provided by the surface distribution of hydrophobic groups. Such an assumption may be most easily tested for isologous interactions of globular subunits [12], i.e., those in which homologous subunits interact through homologous sites. The simulation of the second phase of self-assembly may be performed by minimization of the energy of the interaction between the sites outlined at the first phase.

We have simulated the first phase of self-assembly for 6 different protein monomers participating in known quaternary structures [13–18], this has led to the successful predictions of the most favourable interaction sites. Simulation of the second phase was done for dimeric structures of two proteins and has led to successful prediction of their quaternary structures.

2. Methods

A simplified representation of the side-chains by spheres with their centers in the C^{β}-atoms (C^{α}-atoms for Gly), corresponding approximately to the averaged positions of the side chain centers, is used. This makes the rotations of side-chains and small changes in the main-chain atom positions unimportant for our computations. The coordinates of the C^{β}-atoms were taken from [13–17].

During simulation of the first phase of the self-assembly the most favourable contact regions implicated are those in which maximum gain of hydrophobic energy occurs on intersubunit contact. This gain is proportional to the area of the dehydrated surface and to its mean hydrophobicity per $Å^2$, the area being dependent on the curvature of the smoothed surface. To roughly account for this dependence the contacting surfaces are approximated by ellipsoids of revolution with the centers in the centers of gravity of the monomers (equal weights are assigned to all simplified side-chains). The isologous interaction of monomers is described as interaction of equal ellipsoids by identical sites (fig.1a). For the approximation the subunit was oriented as in fig.1b, the line connecting the subunit centers being on the Z-axis, and the subunit center being in XY-plane. The half-length of the ellipsoid axis of symmetry is taken to be equal to the maximal value of the Z-coordinate of the groups in a cylinder of radius of 7 Å constructed around the Z-axis. The half-length of another ellipsoid

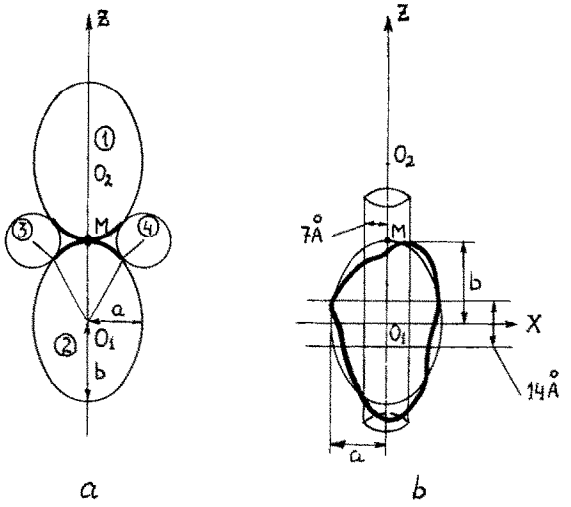

Fig.1. (a) Isologous interaction of two ellipsoids. (1,2) Ellipsoidal monomers; (3,4) water molecules. The dehydrated regions are shown by bold lines. (b) Approximation of the monomer by the ellipsoid of revolution. (0_1, 0_2) Centers of subunits 1 and 2; (M) contact center; (a,b) ellipsoid axes. The subunit is drawn in bold line.

axis is taken equal to the maximum value of $\sqrt{x^2 + y^2}$ for the groups in the layer 14 Å high parallel to the XY-plane (fig.1b). The region dehydrated on contact of the ellipsoids is assumed to be the part of their surface inaccessible for contact with the spheres approximating water molecules (fig.1a). All the hydrophobic groups (Ala, Cys, Met, Val, Pro, Leu, Ile, Phe, Tyr, Trp) for which the projections of their centers from the center of the ellipsoid onto its surface are in the dehydrated regions and are assumed to be dehydrated on intersubunit contact. Hydrophobicity was estimated according to [19]. The accessibility of side chains to water in monomers was estimated with the modified algorithm [20]. The search for the most favourable contact regions was performed by randomly changing the position of the contact region center (point M in fig.1), beginning from the centers of all surface side chains. Each position of the M-point determines the ellipsoid axes and thus the dehydrated region. To describe the predicted contact regions their hydrophobicities and positions of their centers are used. The center of experimentally known contact region is assumed to be at the intersection of the ellipsoid surface and the axis of the cone of revolution having minimum top angle and including all the groups of the contact region. The cone top is in the ellipsoid center.

In the simulation of the second phase of self-assembly the following interactions are taken into account: van der Waals, hydrophobic, hydrogen bonds. The first is taken into account as in [21] with the interaction constant 0.2 equal for all pairs of groups and the second-by a method similar to [22]. The energy of dehydration of polar groups (except Tyr) was taken to be +4 kcal/mol. It was assumed that the formation of H-bond completely neutralizes unfavourability of a polar group dehydration. The energy of these bonds was estimated by the expression $Eij = 1/(R^4 + 1)$, where R is the distance between the centers of the groups i and j minus the sum of their radii. Account was taken that the number of hydrophobic or H-bonds is limited for any side chain. Minimization was carried out by the simplex method [23] from ~50 initial points differing by mutual displacements of the centers of predicted contact regions by 4 Å along coordinate axes and by rotations around the line connecting the subunit centers. The energy of electrostatic interactions between charged groups was calculated for the most favourable structures found by the minimization. The energy was taken to be ± 3 kcal/mol for charged groups in contact and decreased as $1/(R + 3)$.

3. Results and discussion

The results of our search for the most favourable regions of isologous contact are given in table 1. It shows that the hydrophobicity of the regions predicted as the most favourable for the intersubunit interactions roughly reflects the tendencies of the subunits to dimerize. The value obtained for the monomeric protein myoglobin [24] is 2–6 kcal/(mol X monomer) smaller than for other subunits studied. To compare the predicted contact regions with those from the X-ray studies [13–18] the distances between their centers are used. Table 1 shows that 1–5 comparably favourable regions of isologous contact were predicted for each subunit. The centers of one or more predicted regions deviate from the centers of the experimentally localized regions by 1–6.5 Å. This is within the limits of accuracy of our method

Table 1
The most favourable contact regions

Subunit[a]	No. of the predicted contact region	Hydrophobicity (kcal/mol)	Center of the predicted region (Å)			Distance between the centers of the predicted region and experimentally localized region 1[c] (Å)	Distance between the centers of the predicted region and experimentally localized region 2[c] (Å)
			X	*Y*	*Z*[b]		
HA	I	8.5	−5.0	3.3	−11.2	14.6	13.7
	II	8.8	3.7	11.2	−7.7	4.7	19.0
	II	6.9	11.2	7.8	−11.4	6.3	29.7
HB	I	10.3	−3.6	−11.0	5.5	6.2	12.3
	II	8.6	2.0	−12.7	6.0	2.8	17.0
	III	8.4	−3.0	−17.3	3.6	12.6	13.4
TPI	I	9.7	2.1	18.7	1.1	4.5	
	II	7.8	3.9	21.9	−4.6	3.1	
	III	8.6	1.8	19.3	11.2	15.3	
CONC	I	9.2	−7.4	5.5	17.2	20.1	35.6
	II	7.8	−2.1	6.9	−13.0	15.5	4.6
	III	7.9	−13.1	15.0	3.5	3.5	26.4
	IV	8.3	−1.3	4.2	−13.4	17.2	3.0
	V	7.3	−16.2	7.0	5.7	11.1	27.3
TAC	I	9.2	1.0	7.6	−12.5	3.0	
IN	I	7.6	−6.6	−4.3	−7.2	17.7	4.1
	II	7.0	−1.0	−9.7	−5.3	16.7	4.2
	III	6.1	9.8	2.7	−3.3	1.2	17.5
MB	I	4.6	0.1	−11.6	−6.5		
	II	4.2	0.2	−15.0	14.8		
	III	4.0	15.7	−6.4	5.3		
	IV	4.3	6.6	−12.0	−6.3		
	V	3.7	−11.3	16.3	−1.6		

[a] HA designates the α-chain of horse haemoglobin; HB, haemoglobin β-chain [13]; TPI, triose phosphate isomerase [14,18]; CONC, concanavalin [15]; TAC, α-chymotrypsin [16]; IN, insulin [17]; MB, myoglobin [24]

[b] Coordinates are given in coordinate system with the origin in the subunit center and the axes parallel to those used by the authors of the experimental work

[c] Experimentally localized region 1: for HA and HB = contact $\alpha_1\beta_1$ [13]; for CONC = contact between subunits 1 and 2 [15]; for TPI and TAC = contact between monomers in dimer; for IN = contact between monomers [17]. Experimentally localized region 2: for HA and HB = contact $\alpha_1\beta_1$ [13]; for CONC = contact between subunits 1 and 3 [15]; for IN = contact dimer–dimer [17]

For each of the predicted contact regions the table gives the designation of the subunit, the designation of the predicted contact region (by roman figure), – the hydrophobicity of this region, – the coordinates of its center, – the deviations of this center from the centers of each of the experimentally localized regions (these deviations are the measure of the success of the prediction)

being determined by the point representation of hydrophobic groups and by the distance between the centers of two contacting groups (~7 Å). These results show that the surface distribution of hydrophobic groups does provide the rough recognition allowing the prediction of experimentally known regions. The extra regions predicted may be due to deviations of the subunit shape from the ellipsoidal one. For haemoglobin subunits only $\alpha_1\beta_1$ contact is predicted. This may mean that the $\alpha_1\beta_1$ recognition takes place only after dimerization and that for haemoglobin the surface distribution of hydrophobic groups determines the assembly pathway.

We have simulated the second phase of self-assembly for the $\alpha\beta$ dimer of haemoglobin and for the α-chymotrypsin dimer. The results are given in table 2. Three most favourable structures are predicted for the haemoglobin dimer and two for the α-chymotrypsin dimer. For structure 1 predicted for the haemoglobin $\alpha\beta$ dimer the root mean square deviation [21] of the side chain center positions from those known from X-ray analysis is 2 Å, the corresponding value for structure 1 predicted for the α-chymotrypsin dimer is 3 Å. Thus these structures are very similar to the experimentally known ones [13,16]. The last column of table 2 shows that for these structures electrostatic interactions between charged groups are more favourable than for the alternative ones.

Our calculations have shown that the inclusion of H-bonds and their compensating effect into the calculation is important as this decreases the number of favourable structures ~4-fold. But the result was still not unique and the additional inclusion of charged group interactions led to the selection of the unique structure only for α-chymotrypsin. It is probable that electrostatic interactions may also provide an additional type of rough recognition operating at longer distances. The study of this possibility as well as computations for a few other proteins are in progress now. The results described in this paper are only the first steps in computation of protein quaternary structures.

Acknowledgement

We thank Professor G. R. Ivanitsky for support and useful discussions.

References

[1] Wright, H. T., Kraut, J. and Wilcox, P. E. (1967) J. Mol. Biol. 37, 363–366.
[2] Huber, R., Kukla, D., Bode, W., Schwager, P., Bartels, K., Deisenhofer, J. and Steigemann, W. (1974) J. Mol. Biol. 89, 73–101.
[3] Beychok, S., Tyuma, I., Benesch, R. E. and Benesch, R. (1967) J. Biol. Chem. 242, 2460–2462.
[4] Graham, T. and Butler, P. J. (1978) Eur. J. Biochem. 83, 523–528.

Table 2
The most favourable dimeric structures

Protein	Predicted structure	Energy obtained by minimization (kcal/mol)				RMS[a] (Å)	Energy of charge interactions[b] (kcal/mol)
		Total	Contributions				
			Hydrophobic	Van der Waals	H-bonds		
Haemo-	1	−31.5	−10.5	−15.0	−6.0	2.3	−0.7
globin	2	−31.5	−12.2	−13.6	−5.7	23.4	+2.8
	3	−32.3	−10.2	−15.1	−7.0	28.0	+0.6
Chymo-	1	−33.6	−12.8	−12.5	−8.3	3.0	+7.5
trypsin	2	−31.7	−12.9	−10.8	−8.0	18.2	+17.3

[a] RMS = root mean square deviation of the side chain centers positions from those known from the X-ray analysis; RMS is calculated only for 1 subunit, another being in the native position
[b] The energy of charge interactions was not included into the total energy which was minimized; it was calculated for the best structures obtained by the minimization

[5] Kauzmann, W. (1959) Advan. Prot. Chem. 14, 1–63.
[6] Chothia, C. and Janin, J. (1975) Nature 256, 705–708.
[7] Lazdunski, M., Vinsent, J.-P. and Schweitz, H. (1974) Bayer-Symp. V 'Proteinase Inhibitors', pp. 420–431.
[8] Turner, R., Liener, I. E. and Lovrien, R. E. (1975) Biochemistry 14, 275–282.
[9] Blow, D. M. and Smith, J. M. (1975) Phil. Trans. Roy. Soc. Lond. B 272, 87–97.
[10] Valdes, R. and Ackers, G. K. (1977) J. Biol. Chem. 252, 74–81.
[11] Huet, M. and Claverie, J. M. (1978) Biochemistry 17, 236–241.
[12] Monod, J., Wyman, J. and Changeux, J.-P. (1965) J. Mol. Biol. 12, 88–118.
[13] Perutz, M. F., Muirhead, H., Cox, J. M. and Goaman, L. C. G. (1968) Nature 219, 131–139.
[14] Banner, D. W., Bloomer, A. C., Petsko, G. A., Phillips, D. C. and Wilson, I. A. (1976) Biochem. Biophys. Res. Commun. 72, 146–155.
[15] Reeke, G. N., Becker, J. W. and Edelman, G. M. (1975) J. Biol. Chem. 250, 1525–1547.
[16] Birktoft, J. J. and Blow, D. M. (1972) J. Mol. Biol. 68, 187–240.
[17] Blundell, T., Dodson, G., Hodgkin, D. and Mercola, D. (1972) Advan. Prot. Chem. 26, 280–402.
[18] Banner, D. W., Bloomer, A. C., Petsko, G. A., Phillips, D. C., Pogson, C. I., Wilson, I. A., Corraan, P. H., Furth, A. J., Milman, J. D., Offord, R. E., Proddle, J. D. and Waley, S. G. (1975) Nature 255, 609–614.
[19] Tanford, C. (1962) J. Am. Chem. Soc. 84, 4240–4247.
[20] Shake, A. and Rupley, J. A. (1973) J. Mol. Biol. 79, 351–371.
[21] Levitt, M. (1976) J. Mol. Biol. 104, 59–107.
[22] Gibson, K. D. and Scheraga, H. A. (1967) Proc. Natl. Acad. Sci. USA 58, 420–427.
[23] Nelder, J. A. and Mead, R. (1965) Comput. J. 7, 308–315.
[24] Watson, H. C. (1968) Progr. Stereochem. 4, 299–333.