LONG-RANGE ORDER IN GLOBULAR PROTEINS

B. A. FEDOROV

Institute of Protein Research, Academy of Sciences of the USSR, 142292 Poustchino, Moscow Region, USSR

Received 13 November 1975

1. Introduction

X-ray diffraction analysis of proteins gives abundant statistical data for studying particular and general regularities of their internal structure. One of such general regularities, which seems to apply to globular proteins is described in this paper. This is a marked periodicity in the alternation of denser and less dense regions of the polypeptide chain within the protein globule. This periodicity which is based on the tight packing of the protein molecule chains extends over long and short distances thus covering the whole globule.

To substantiate this conclusion, the function of the distribution of distances between all the non-hydrogen atoms of the protein molecule must be considered. However, to calculate this function, it is necessary to introduce two principal modifications which remove subsidiary effects from the distribution function and increase its ability to resolve internal inhomogeneities in the particle. Firstly the contrast between the dense and less dense regions of the polypeptide chain must be increased to a maximum and secondly, the distribution function should be made independent of the influence of dimensions and shape of the protein molecule. If these points are not taken into account the characteristic features of the distribution function can completely disappear, especially for middle and large distances [1,2].

2. Method of calculation

To fulfil the first requirement, i.e. to achieve a maximum in contrast, it is not the value $\bar{\rho}(\vec{r})$ that must be considered as the electron density of the particle at a point $\vec{r}(r,\vartheta,\varphi)$ but the deviation $\rho(\vec{r})-\bar{\rho}$ of this value from the average protein value. The three-dimensional distribution function (the three-dimensional function of interatomic distances) can be described as

$$D(\vec{r}) = \int_V (\rho(\vec{r}'+\vec{r})-\bar{\rho}(\vec{r}'+\vec{r}))(\rho(\vec{r}')-\bar{\rho}(\vec{r}'))dV_{\vec{r}'} \quad (1)$$

and the one-dimensional distribution function as

$$D(r) = \int_0^{2\pi} d\varphi \int_0^{\pi} D(r,\vartheta,\varphi)\,\mathrm{Sin}\,\vartheta\,d\,\vartheta. \quad (2)$$

Here V is the molecule volume and

$$\bar{\rho}(\vec{r}) = \begin{cases} \bar{\rho} \text{ if } \vec{r} \text{ inside V,} \\ 0 \text{ if } \vec{r} \text{ outside V.} \end{cases}$$

The second requirement is reduced to normalizing the obtained one-dimensional distribution function D(r) by the correlation function for the homogeneous particle having the same dimensions and shape as the considered inhomogeneous particle. This correlation function is

$$D_h(r) = \frac{1}{4\pi}\int_0^{2\pi} d\varphi \int_0^{\pi} D_h(r,\vartheta,\varphi)\,\mathrm{Sin}\,\vartheta\,d\,\vartheta, \quad (3)$$

where

$$D_h(\vec{r}) = \frac{1}{V\bar{\rho}^2}\int_V \bar{\rho}(\vec{r}'+\vec{r})\bar{\rho}(\vec{r}')dV_{\vec{r}'}, \quad (4)$$

It can be seen that the function D(r) depends not

only on the character of distribution of internal inhomogeneities but also on geometric external parameters of the protein globule. Thus, for example, when r is close to maximum dimensions of the particle the number of atom-pairs, whose atoms are separated by the distance r, considerably decreases and D(r) inclines to zero practically independently of the character of distribution of inhomogeneities at these distances. However, if we take into account the geometric meaning [3] of the correlation function $D_h(r)$ i.e., the probability of finding the particle at a distance r from a random point of the same particle, the dependence of this probability against r is eliminated in the function

$$D_{inh}(r) = \frac{D(r)}{D_h(r)} \qquad (5)$$

and thus the dependence of the distribution function on the geometric parameters of the protein molecule is completely eliminated.

Formulae (1)–(5) serve as the basis for calculation of distribution of inhomogeneities for molecules with known atomic coordinates. Consideration of the value $\bar{\rho}$ in the integral (1) can be done correctly enough by methods developed earlier, taking into account the influence of solvent on X-ray diffuse scattering by proteins in solution [4]. At the same time a simple way of calculating integral (1) is quite efficient when the weighting function (the effective atomic scattering factor) for each atom is taken as the difference

$$f_i - k\bar{\rho}V_i$$

where f_i is the number of electrons of the i-th atom and the hydrogen atoms adjoined to it (their atomic scattering factor at zero angle of scattering), V_i is their total van der Waals volume, $k=V/\sum_i V_i$ is the coefficient compensating for the discrepancy between the molecule volume and the sum of the van der Waals volumes of its atoms. The correlation function $D_h(r)$ can be calculated in an analogous way: the integral (4) should be substituted by a summation on the basis of the same atoms, the value $k\bar{\rho}V_i$ being taken as the effective atomic scattering factor of the i-th atom. This approximation for calculation of D(r) and $D_h(r)$ is quite satisfactory at all distances except the shortest (about 0.1–0.7 nm) where a noticeable influence of the discrete localization of atoms distorts $D_h(r)$ and partly also D(r).

3. Results and discussion

Calculation of the distribution function of inhomogeneities $D_{inh}(r)$ was performed according to formulae (1)–(5) taking into account the above approximations for a number of globular proteins with a known three-dimensional structure: sperm-whale myoglobin [5], lamprey globin [6], hen egg-white lysozyme [7] and bovine α-chymotrypsin [8] (references indicate origin of coordinates). The same value $\bar{\rho}=0.425$ el.A^{-3}, which is the average electron density of the proteins and $k=1.4$ were taken for all the four proteins. Fig.1 represents the results of calculation. Significant periodicity covering small and large distances is observed for all the proteins (the maximum connected with the van der Waals distance of 0.45 nm is not displayed because of the calculation error mentioned above). The 1 nm maximum corresponding to the strong maximum observed at ~ 9° (for CuK_α-radiation) on the X-ray diffuse scattering curve by globular proteins was interpreted earlier [9] as a packing maximum for the polypeptide chains neighbouring in space. At the same time the presence of the 2 nm maximum and moreover of the 3 and 4 nm maxima was completely suppressed

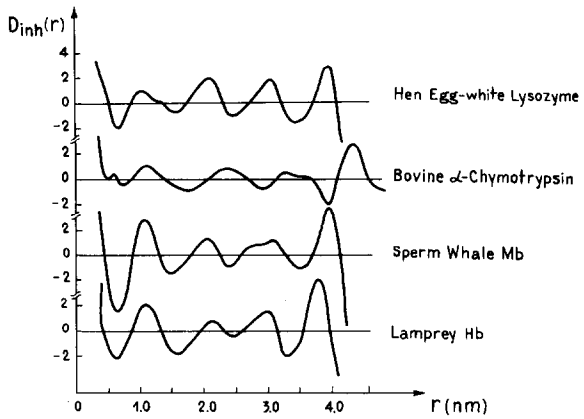

Fig.1. Inhomogeneity function distribution corrected for molecular shape calculated from coordinates of atoms for four globular proteins.

on the curve of X-ray scattering by proteins due to a sharp decrease of intensity in the small angle range.

It is important to note (fig.1) that long range order covers the largest distances in the molecule (4 nm), which practically corresponds to the maximum dimension of these proteins. This demonstrates a higher degree of ordering (at least for the indicated four proteins) in the general three-dimensional organization of globular proteins than has been known earlier.

Acknowledgements

The author is grateful to Professor O. B. Ptitsyn for valuable discussions and A. I. Denesyuk for the help in calculations.

References

[1] Watson, H. C. (1967) Small-Angle X-ray Scattering, (Brumberger, H., ed.) pp. 267–275, Gordon and Breach, New York.
[2] Krigbaum, W. R. and Kügler, F. R. (1970) Biochemistry 9, 1216–1223.
[3] Guinier, A. and Fournet, G. (1955) Small-Angle Scattering of X-rays, Willey & Sons, Inc., New York.
[4] Fedorov, B. A., Ptitsyn, O. B. and Voronin, L. A. (1974) J. Appl. Cryst. 7, 181–186.
[5] Watson, H. C. (1969) Progr. Stereochem. 4, 299–333.
[6] Hendrickson, W. A., personal communication.
[7] Imoto, T., Johnson, L. N., North, A. C. T., Phillips, D. C. and Rupley, J. A. (1972) in: The Enzymes, 3rd Edn. (Boyer, P., ed.) Vol. 7, pp. 665–868, Academic Press, New York and London.
[8] Birktoft, J. J. and Blow, D. M. (1972) J. Mol. Biol. 68, 187–240.
[9] Echols, G. E. and Anderegg, J. W. (1960) J. Am. Chem. Soc. 82, 5085–5092.