Biophys Struct Mech (1983) 10: 31–45

Biophysics of Structure and Mechanism

Proteins Main-Chain Atomic Displacements and Density of Stabilizing Interactions

B. Gavish

Department of Physics, University of Illinois at Urbana-Champaign, 1110 West Green Street, Urbana, Illinois 61801, USA

Abstract. The assumption that interactions like hydrogen bonds, that establish the secondary structure of proteins, modulate the local flexibility of the polypeptide chain suggests a phenomenological relation between σ^2- the X-ray determined variance of the thermally-distributed location of a main-chain atom and the value of ϱ – a properly-defined linear density of "stabilizing interactions" at that location. The functional relation $\sigma^2 \alpha 1/\varrho$ is verified from the data of lysozyme. ϱ is constructed from first principles after assuming that the thermal motions of an unstabilized polypeptide chain resemble those of a random chain. Taking the locations and the identity of the stabilizing interactions from literature, the predicted $1/\varrho$ for lysozyme and metmyoglobin is compared with the observed σ^2 along the proteins main-chain. The satisfactory results are discussed in the light of the possible role that hydrogen bonding plays in determining the equilibrium and the dynamic properties of the main-chain structural fluctuations in proteins, and its modelling by using a simplified mechanistic approach.

Key words: X-rays determined structural fluctuations – Local flexibility – Hydrogen bonds – A random chain

Introduction

Recent X-ray studies of the thermal atomic displacements in proteins (Frauenfelder et al. 1979; Artmyiuk et al. 1979; Sternberg et al. 1979) shed new light on protein structure by converting the traditional "frozen" X-rays structure into a dynamical one. The observed variation of the mean square atomic displacement σ^2 along the main-chain atoms is found to be very nonuniform and possesses the following features: 1) The existence of high-σ^2 regions which correspond to main-chain segments that are not involved in the regular

Present address: Rogoff Department of Biomedical Engineering, The Hebrew University-Hadassah Medical School, P.O. Box 1172, Jerusalem, Israel

secondary structure (Artmyiuk et al. 1979; Sternberg et al. 1979). Similar features were obtained by computer simulations (McCammon and Karplus 1980). 2) σ^2 increases with increasing the distance from the center of the proteins (Artmyiuk et al. 1979; Sternberg et al. 1979). 3) σ^2 at high-σ^2 regions reveals a weak temperature dependence (Frauenfelder et al. 1979) which can be interpreted in terms of "flat" local potentials (Frauenfelder et al. 1979; Gavish 1981). However, with these observations and correlations one is still unable to predict *quantitatively* the variations of σ^2 along the main chain from a given X-ray structure of a protein. It is the purpose of this work to find, from the experimental data, a quantitative relation between σ^2 and a structural parameter, which is a measurable quantity, and to test with this phenomenological relation some assumptions about the "mechanics" of the polypeptide chain in its folded state.

Theory

A Phenomenological Approach

The qualitative correlation between the occurrence of high-σ^2 regions and irregularities in the secondary structure motivates the following idea: In the lack of hydrogen bonding and other interactions that establish the secondary structure, the polypeptide chain is likely to resemble a random chain which is highly flexible. This is due to the rotational degrees of freedom of the peptide bonds. It seems reasonable that a distribution of these interactions along the polypeptide chain leads to spatial modulation of its flexibility. We shall refer hereafter to these interactions as *stabilizing interactions* by means of their proposed role in stabilizing the polypeptide chain against structural fluctuations. An example to such system is illustrated in Fig. 1A.

The next logical step is to assume that there is a local relation between σ^2 and ϱ – a properly defined linear density of stabilizing interactions in the following form:

$$\varrho(r_0) = g[\sigma_0^2/\sigma^2(r_0)]\,, \tag{1}$$

where $r = 1, 2, \ldots, N$ is the residue number, g and σ_0^2 are an unknown function and an adjustable parameter, respectively. Taking $n(r_0)$ to be the number of stabilizing ineractions at residue r_0, $\varrho(r_0)$ should be a smoothed form of $n(r_0)$. Therefore, regardless of its definitions, the following relation should approximately hold:

$$\sum_{r_0=1}^{r} \varrho(r_0) \approx \sum_{r_0=1}^{r} n(r_0) \tag{2}$$

We shall assume for simplicity that *all the included interactions have identical stabilizing power by means of their effect on the fluctuations of the polypeptide chain.* Further justification to this assumption will be given in the Discussion. Eq. (2) is of importance because $n(r_0)$ can be deduced from the X-ray structure

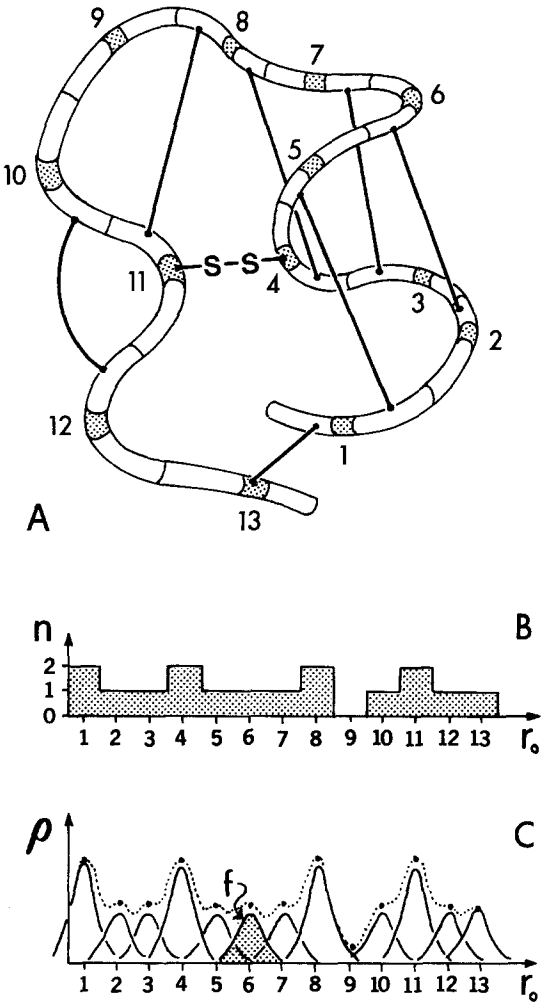

Fig. 1. A A schematic presentation of a polypeptide chain, subject to stabilizing interactions. The units 1–13 denote amino acid residues. The shaded segment in the center of each unit stands for the α-carbon. The thick lines represent hydrogen bonds between different peptide bonds, e.g., 1–5, or between a peptide bond and a side chain, e.g., 1–13. Side chains are not shown. **B** The number of stabilizing interactions per residue $n(r_0)$ versus the residue number r_0 for the "protein" shown in **A**. **C** The construction of a smooth density of stabilizing interactions $\varrho(r_0)$ is shown for a continuous r_0 as a dotted line and is represented at discrete values by filled circles (in arbitrary units). $\varrho(r_0)$ is a sum of $n(r_i)$ given in (**B**) weighted by a distribution function $f(r_i - r_0)$ as given by Eq. (4)

of a protein. $n(r_0)$ of the hypothetical protein shown in Fig. 1A, is shown in Fig. 1B. The simplest form of Eq. (1) is

$$\varrho(r_0) = \sigma_0^2/\sigma^2(r_0) \tag{3a}$$

or alternatively:

$$\sum_{r_0=1}^{r} 1/\sigma^2(r_0) = (1/\sigma_0^2) \sum_{r_0=1}^{r} \varrho(r_0) \approx (1/\sigma_0^2) \sum_{r_0=1}^{r} n(r_0) \,. \tag{3b}$$

Using the data of Artmyiuk et al. (1979) for lysozyme with $n(r_0) = 0{,}1$ denoting the residues, which are involved in intramolecular hydrogen bonds (in the lack of more detailed information), we can test the validity of Eq. (3b). The plot shown in Fig. 2 verifies Eq. (3b) with a correlation coefficient of 0.9966 and

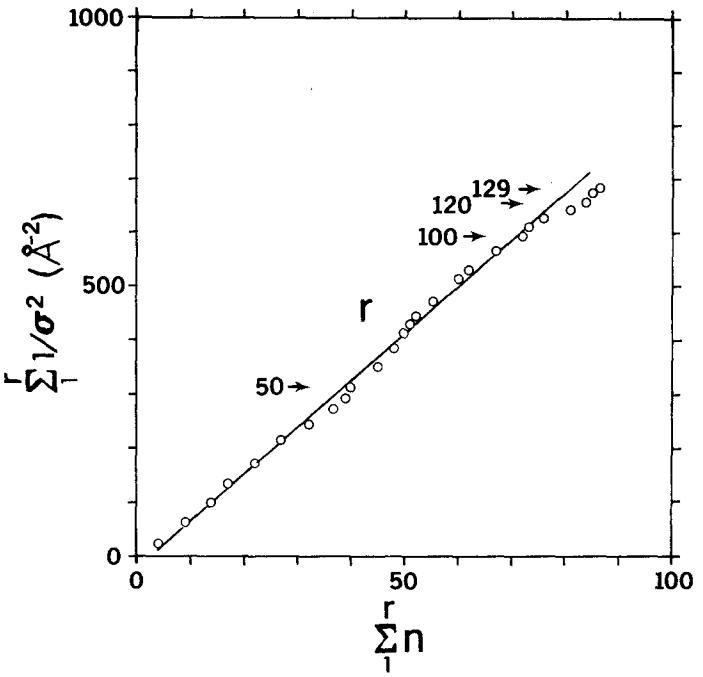

Fig. 2. Testing Eq. (3b) by using the data of lysozyme (Artmyiuk et al. 1979). The open circles represent the summations on r calculated for r = 5, 10, 15, ..., 120, 125, 129. Using linear regression we find σ_0^2 = 0.115 Å^2 with correlation coefficient of 0.9966

$\sigma_0^2 = 0.115$ Å^2. More details are given in the legend. The same test is inapplicable for the data of metmyoglobin because $\sigma^2(r_0)$ reaches values too small with respect to the experimental error. If we neglect the small deviations from linearity Fig. 2 proves that Eq. (3a) is the only possible form of Eq. (1) that fits these data.

Constructing $\varrho(r_0)$

In spite of the above results the plot of Fig. 2 is insensitive to the structure of $\sigma^2(r_0)$ at the peaks, which seem to be the most informative parts of the data. Therefore, we should test the relation

$$\sigma^2(r_0) = \sigma_0^2/\varrho(r_0) \tag{3c}$$

or alternatively

$$\sum_{r_0=1}^{r} \sigma^2(r_0) = \sigma_0^2 \sum_{r_0=1}^{r} 1/\varrho(r_0) \tag{3d}$$

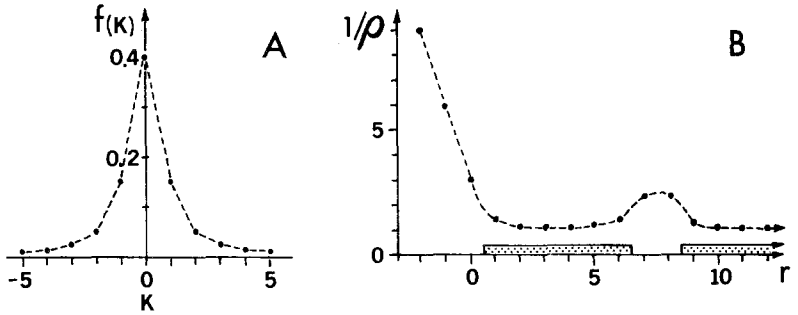

Fig. 3. A A plot of the distribution function $f(k)$ used in constructing $\varrho(r_0)$ (see Eq. 4) derived for the case of a random chain in the appendix (see Eq. A-4). The plot involves $A = 0.3$. **(B)** The calculated $1/\varrho(r)$ for an infinite chain $-\infty < r < +\infty$, stabilized at r values that belong to the shaded area, which extends to $+\infty$. The calculations involved the $f(k)$ function shown in A

which, in contrast to Eq. (3b), is sensitive to the structure of $\varrho(r_0)$. The next step is to try to derive $\varrho(r_0)$ by using further assumptions about the flexibility of the polypeptide chain $\varrho(r_0)$ can be generated to be a smooth form of $n(r_0)$ by using the following expression:

$$\varrho(r_0) = (1/p) \sum_{r_i = -\infty}^{\infty} f(r_i - r_0) n(r_i) = (1/p) \sum_{k = -\infty}^{\infty} f(k) n(r_0 + k) , \tag{4}$$

where $f(k)$ is a distribution function and $n(r_i) = 0, 1, \ldots, p$. We assume that $f(k)$ *is independent of the residue and the type of stabilizing interactions under consideration.* The normalization of $f(k)$ requires:

$$\sum_{k = -\infty}^{\infty} f(k) = 1 . \tag{5}$$

We shall further assume $f(k)$ to be a symmetric function of k. It turns out that for a fully-stabilized chain, i.e., $n(r_i) = p$, $\sigma^2(r_0) = \sigma_0^2$. A hypothetical $\varrho(r_0)$ which has been constructed in this way is illustrated schematically in Fig. 1c.

Returning to our basic considerations we assume that *a polypeptide chain, in the lack of stabilizing interactions, can be reasonably described as a random chain.* It is shown at the appendix that in this case

$$f(k) = \begin{cases} 1 - 2A & k = 0 \\ A/|k|(|k| + 1) & k \neq 0 \end{cases}, \tag{6}$$

where A is an adjustable parameter. $A = \sigma_0^2/U_0^2$ where U_0^2 is the mean-square relative displacement of a unit in a random chain. We therefore expect $A < 1$. $f(k)$ from Eq. (6) is plotted in Fig. 3A. Using Eqs. (3c) and (6), it is shown in the appendix that for an infinite chain, unstabilized merely at a segment of length l, $\sigma^2(0) = U_0^2 l/4$ at the middle of this segment [see Eqs. (A-5) and (A-6) and Fig. 3B]. If the chain has a "free end", i.e., unstabilized for $r < 0$ we obtain $\sigma^2(-l) = U_0^2 l$ [see Eqs. (A-1) and (A-2) and Fig. 3B].

Methods and Results

General Procedure

The prediction of this theory as given by Eqs. (3c), (3d) was tested for lysozyme and metmyoglobin for which $\sigma^2(r_0)$ and $n(r_0)$ are known. The testing procedure involved the following steps:

a) Choosing a value of A and using Eqs. (4) and (6) for a known $n(r_0)$, $\varrho(r_0)$ was calculated.

b) The quantity $\sum_{r_0=1}^{r} \sigma(r_0)$ was plotted versus $\sum_{r_0=1}^{r} 1/\varrho(r_0)$ and σ_0^2 and the correlation coefficient were found by using a linear regression (see Eq. 3d).

c) $1/\varrho(r_0)$ was superimposed on $\sigma^2(r_0)$, taking the above A and σ_0^2 values.

d) Repeating steps (a) to (c) for different values of A, the A-value corresponding to $1/\varrho(r_0)$ curve which showed the closest fitness to $\sigma^2(r_0)$ was found. This value determined σ_0^2.

e) The steps (a) to (d) were repeated for different $n(r_0)$ distributions (for metmyoglobin only) in order to study contributions of different kinds of stabilizing interactions. Furthermore, in order to test the sensitivity of the results to the structure of $f(k)$ a Lorentzian form of $f(k)$ was used, i.e., $f(k) = B\lambda/(\lambda^2 + k^2)$ where B is a normalization constant and λ is an adjustable parameter.

Lysozyme

Contains 129 residues. The values of $\sigma^2(r_0)$ were reproduced from Artmyiuk et al. (1979) as well as $n(r_0)$ in terms of "all or nothing" distribution of residues, that are involved in intramolecular hydrogen bonding (this was done in the lack of more detailed information which was not available). Disulfide bridges were taken into account provided that there was no intramolecular hydrogen bond at that location. Eq. (3d) was found to fit the data for $A = 0.32$ and $\sigma_0^2 = 0.1136$ Å^2 with a correlation coefficient of 0.9939 (Fig. 4). More details are given at the legend of Fig. 4. Figure 5 shows a reasonable agreement between the calculated $1/\varrho(r_0)$ values and the observed $\sigma^2(r_0)$ ones. In some regions (40–55 and 100–110) the predicted values show resemblence with the fine structure of $\sigma^2(r_0)$. The predicted peak around $r_0 = 70$ is somewhat shifted with respect to the observed one but still reveals a striking similarity with the data. However, $1/\varrho(r_0)$ deviates considerably from $\sigma^2(r_0)$ for $r > 110$. A reasonable agreement between $1/\varrho(r_0)$ and $\sigma^2(r_0)$ was obtained as well upon varying A by $\pm$ 3% around its best-fit value. The contribution of the disulfide bond, which was taken into account at $r_0 = 76$ and 127, can be shown to be significant in improving the fitness of $1/\varrho(r_0)$ to $\sigma^2(r_0)$. Using a Lorentzian form of $f(k)$ (see Procedure), a general agreement is obtained between $1/\varrho(r_0)$ and $\sigma^2(r_0)$ for $\lambda = 2$, $B = 0.743$. σ_0^2 and the correlation coefficient values are very close to the above ones. However, $1/\varrho(r_0)$ displays a smooth behavior and does not possess the fine structure of $\sigma^2(r_0)$ as obtained by using the Eq. (6) form of $f(k)$.

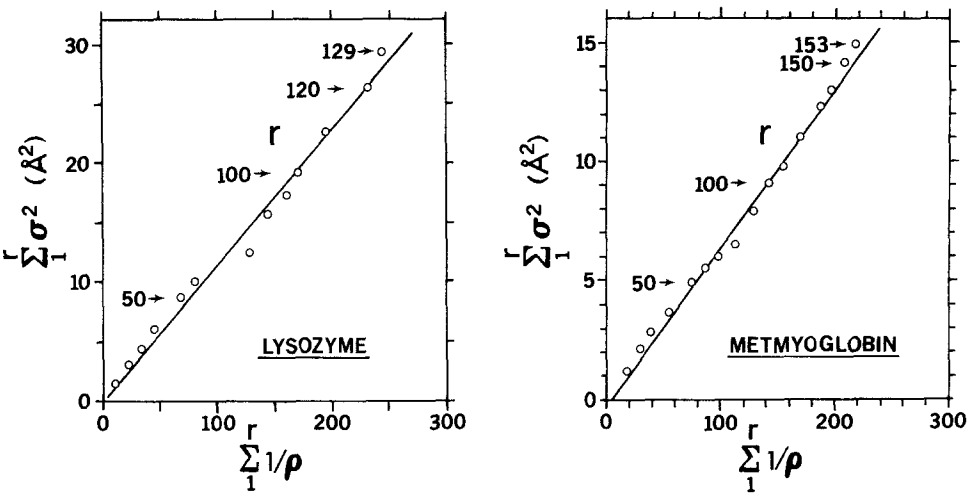

Fig. 4. Shows the testing of Eq. (3d) for the $\varrho(r_0)$ given by Eq. (4) and the $f(k)$ given by Eq. (6). The references from which the data of $\sigma^2(r_0)$ and $n(r_0)$ were taken are mentioned in the legend of Fig. 5. *Lysozyme:* The best-fit A-value was 0.32. The data points represent the summations for r = 10, 20, . . ., 120, 129. The r-values used for the linear regression were r = 5, 10, . . ., 125, 129. This yielded σ_0^2 = 0.1136 Å^2 with correlation coefficient of 0.9939 for the best-fit line shown. *Metmyoglobin:* The best-fit A-value was 0.30. The data correspond to r = 10, 20, . . ., 150, 153. Those used for linear regression involved r = 5, 10, . . ., 150, 153. This gave σ_0^2 = 0.067 Å^2 with correlation coefficient of 0.9964. The plotted results correspond to $n(r_0)$ which included the combined (a), (b), and (c) types of stabilizing interactions (see Methods and Results).

Metmyoglobin

Contains 153 residues. The values of $\sigma^2(r_0)$ at 300 K were supplied by G. A. Petsko and used with permission after being published in Frauenfelder et al. (1979). The construction of $n(r_0)$ was based on the 2.0 Å resolution refined structure of this protein (Takano 1977). The following stabilizing interactions have been considered: a) Hydrogen bonds of the type $\succ$CO$-$HN$\prec$ where *both* N and O belong to peptide bonds. b) Similar to a) but only one of N or O belongs to a peptide bond, i.e., the other one belongs to a side chain or to the heme group. This type of interaction is marked by R in Fig. 5. We included in this category also $\succ$CO$-$H$_2$O$-$HN$\prec$ where N *or* O or *both* belong to peptide bonds. This case is denoted by W in Fig. 5. c) Salt bridges indicated by S in Fig. 5. Since the rotational freedom of the peptide bond is presumed to dominate $\sigma^2(r_0)$, type (b) interaction is likely to be less restrictive than type (a) due to the additional degrees of freedom contributed by the residue involved (Janin et al. 1978) or by the mediation of water molecules (Zundel 1976). Type (c) seems to be less restrictive than (b) due to the large flexibility of the combined side chains (one of them is known to be lysine or arginine (Takano 1977). However, its proposed stabilizing effect in proteins (Perutz 1978) and its probable analogy to the disulide bonds in lysozyme, which do not occur in metmyoglogin, justifies its inclusion. Since the presence of two interactions of type (a) per residue

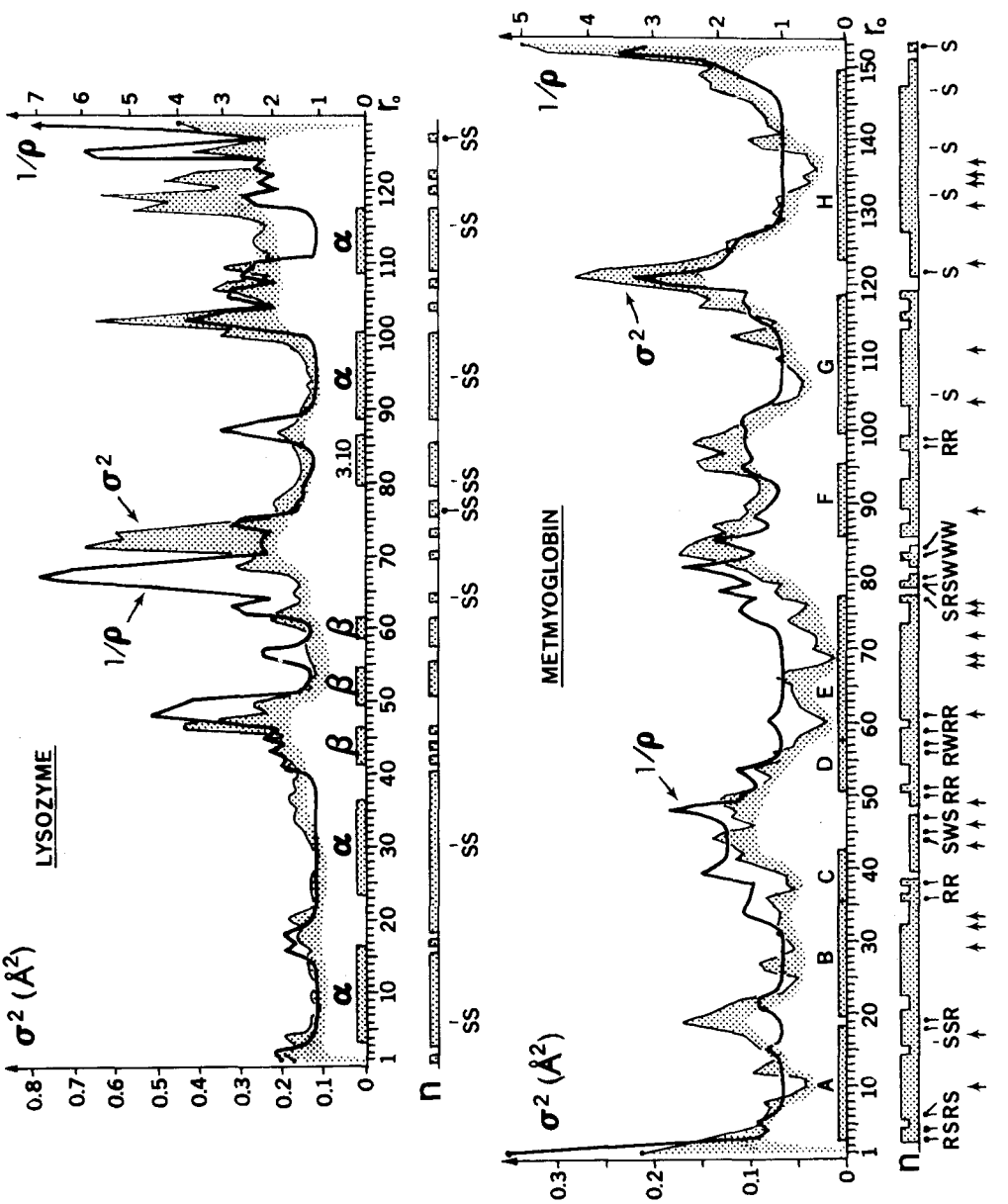

Fig. 5. Shows a comparison between the X-ray values of $\sigma^2(r_0)$ and the calculated $1/\varrho(r_0)$ where r_0 is the residue number. *Lysozyme:* $\sigma^2(r_0)$ and $n(r_0)$ were reproduced from Artmyiuk et al. (1979) with a permission (copyright 1972, McMillan Journal Ltd.). $n(r_0) = 0{,}1$ corresponds to intramolecular hydrogen bonds. SS denotes disulfide bridges, only two of those were taken into account in calculating $n(r_0)$. The locations of the helices α and 3.10 and the β pleated-sheets are marked (after Artmyiuk et al. 1979). The missing point at $r_0 = 129$ correspond to $1/\varrho(r_0) = 8.6$. The correlation coefficient for Eq. (3c) is 0.42 (see Discussion). *Metmyoglobin:* The data of $\sigma^2(r_0)$ at 300 K were supplied by G. A. Petsko [similar data at 250 K are found in Frauenfelder et al. (1979)]. Used with permission; Copyright 1979, McMillan Journal Ltd.]. $n(r_0)$ was taken from Takano (1977) and included, in the plotted case, the following stabilizing interactions: (a) hydrogen bonding between two peptide bonds; (b) hydrogen bonding between a peptide bond and a side chain (R) and a hydrogen bond, similar to the previous ones but through the mediation of water molecule (W); (c) salt bridges (S) that only a fraction of them was taken into account as marked. The arrows represent the locations of the residues involved in hydrophobic interactions. The stripes A to H stand for helical regions (after Frauenfelder, et al. 1979). The correlation coefficient for Eq. (3c) is 0.63 (see Discussion)

characterizes the most stable configuration of the secondary structure, it seems to be sufficient to consider $p = 2$ as the upper limit of $n(r_0)$, i.e., $n(r_0) = 0, 1, 2$. In spite of the proposed differences in the stabilizing power of interactions (a) to (c) we shall assign them an equal weight. Eqs. (3c) and (3d) were tested separately for (a), combined (a) and (b) and combined (a), and (c). 234 stabilizing interactions were considered; 208 out of which were of (a) type, 18 of (b) type (13 of R and 5 of W), and 8 of (c) type. The results for the combined (a), (b), and (c) case are shown in Figs. 4 and 5. It should be noticed that only part of the salt bridges were included due to the restriction $n(r_0) \leq 2$. The best-fit values of the parameters are $A = 0.30$, $\sigma_0^2 = 0.067$ Å^2 (with correlation coefficient of 0.9964). A reasonable fitness between $1/\varrho(r_0)$ and $\sigma^2(r_0)$ is still obtained upon varying A by 10% around its best-fit value; details are given in the legends. The results for $n(r_0)$ based on case (a) or the combined (a) and (b) cases are not shown for simplicity. The general peaks structure is obtained upon the inclusion of merely (a) case. The inclusion of (b) and later (c) cases leads to significant improvement in the local fitness of $1/\varrho(r_0)$ at the peaks regions of $\sigma^2(r_0)$. σ_0^2 slightly decreases upon the inclusion of more interactions. Figure 5 shows the location of residues which are involved in hydrophobic interactions (Takano 1977) (see arrows). There is a positive correlation between the occurrence of clusters of such residues and dips in $\sigma^2(r_0)$ which are not predicted by $1/\varrho(r_0)$ in its present form (see regions 67–77 and 131–139). However, inclusion of *all* the hydrophobic interactions upon using the same procedure as before leads to a significant decrease in the fitness of $1/\varrho(r_0)$ to $\sigma^2(r_0)$. As in the case of lysozyme, a Lorentzian form of $f(k)$ has been applied to the data of the combined (a), (b), and (c) case. $1/\varrho(r_0)$ revealed in this case the general structure of $\sigma^2(r_0)$. However, $1/\varrho(r_0)$ appeared to be rather smooth and lacked the fine structure, which had be obtained by using $f(k)$ given by Eq. (6). The values of the best-fit Lorentzian parameters where $\lambda = 2$ and $B = 0.742$. σ_0^2 and the correlation coefficient values were similar to the above ones.

Discussion

The presented theory assigns to the interactions that stabilize the secondary structure of the protein the role of modulating the local flexibility of the polypeptide chain. A proportionality relation between the reciprocal value of the mean square displacement $\sigma^2(r_0)$ at residue r_0 and the density $\varrho(r_0)$ of the above stabilizing interactions, mainly intramolecular hydrogen bonds (Fig. 1), can be verified from the data of the lysozyme (Fig. 2). $\varrho(r_0)$ can be constructed from first principles by assuming that the unstabilized polypeptide chain possesses random chain features and by attributing to different stabilizing interactions equal weight in determining $\varrho(r_0)$. A proportionality relation between $\sigma^2(r_0)$ and $1/\varrho(r_0)$ was found to be in a good correlation with experimental data for lysozyme and metmyoglobin (Figs. 4 and 5). The results strongly support the idea that the local rigidity of the polypeptide chain can be mainly attributed to hydrogen bonding. Hereby, the discussion will be focused on specific problems.

Similarities and Dissimilarities Between $1/\varrho(r_0)$ and $\sigma^2(r_0)$ in Lysozyme and Metmyoglobin

The location of the $1/\varrho(r_0)$ peaks mostly coincides with those of $\sigma^2(r_0)$. A serious deviation occurs in lysozyme around $r_0 = 70$. Looking at the "fine structure" of the peaks we find that $1/\varrho(r_0)$ possesses the "jurky" character of $\sigma^2(r_0)$ at the peaks regions with a striking similarity in most of the cases. However, the $1/\varrho(r_0)$ sub-peaks have different heights than that of $\sigma^2(r_0)$, e.g., $r_0 = 85-90$ and $115-129$ in lysozyme. Serious disagreements between $1/\varrho(r_0)$ and $\sigma^2(r_0)$ occur near the end of the chain in lysozyme and for both ends in metmyoglobin. It is hard to see a simple explanation to the above deviations. The lack of the observed peak around $r_0 = 19$ in $1/\varrho(r_0)$ of myoglobin is probably due to the fact that the backbone sharply turns there and form an external loop which is favorable for large local fluctuations. The negative deviations of $\sigma^2(r_0)$ from $1/\varrho(r_0)$ in metmyoglobin for $r_0 = 68-78$ and $132-138$ can be attributed to the relative increase in the density of residues which are involved in hydrophobic pockets. An important similarity in the results of our analysis for these two proteins is the close resemblence between the best fit A (or B) values which supports our assumption about the common origin of the $\sigma^2(r_0)$ modulation by the stabilizing interactions.

On the Random Nature of the Unstabilized Segments of the Polypeptide Chain

Local flexibility can be related to the curvature of the potential function at that location (Gavish 1978). Thus, in the lack of harmonic restoring forces, a random chain can be described in terms of motion in "flat" potential wells. σ^2 in a square-well potential can be shown to be temperature independent (Gavish 1981). *The weak temperature dependence observed at the high-σ^2 regions in metmyoglobin (Frauenfelder et al. 1979) strongly supports this picture of motion in non-harmonic local potentials (Frauenfelder et al. 1979; Gavish 1981), which are suggested here to stem from the random nature of the unstabilized polypeptide chain.* It is of interest to note that the low-temperature density of vibrational states of the main-chain atoms in proteins has been found to fit the model of self-avoiding random walk (Stapleton et al. 1981). This is compatible with our picture.

We may conclude that since the energy barriers for the motion of the unstabilized polypeptide chain are much smaller than the barriers which determine the strength of the stabilizing interactions, it is not surprising to find that our model, in which the same weight in determining $\varrho(r_0)$ is assigned to interactions having different stabilizing power, e.g., (a), (b), and (c) in metmyoglobin, fits the data reasonably well. It is of interest to mention that using the relation $A = \sigma_0^2/U_0^2$ (see Appendix), the relative mean square displacement of a single unit in a random polypeptide chain-U_0^2 is found to be 0.36 Å^2 in lysozyme and 0.22 Å^2 in metmyoglobin.

Can $\sigma^2(r_0)$ be Related to the Thermal Stability of the Proteins?

$\sigma^2(r_0)$ is the variance of the equilibrium position distribution function at r_0, which involves the local potential through the Boltzmann's factor. It can be shown that most of the contribution to $\sigma^2(r_0)$ comes from a "slice" of thermal energy $\lesssim 1$ Kcal/mole measured from the bottom of the local potential well. However, the local stability in a given environment is related to the depth of the potential well, which has no appreciable effect on the distribution function that determines $\sigma^2(r_0)$*!. Thus, the value of* $\sigma^2(r_0)$ *is unlikely to be related to the local stability of the protein structure.* Most of the stabilizing interactions which were taken into account here are main-chain−main-chain interactions which do not contribute to the stability of the tertiary structure. The later is likely to originate mainly from hydrophobic interactions (Kuntz 1972) which were demonstrated here not to contribute appreciably to $\sigma^2(r_0)$. Some support for the argument that mobility of the polypeptide chain is not necessarily related to the thermal stability of the protein comes from hydrogen exchange measurements; it was recently found (Hilton et al. 1981) that the observed hydrogen exchange rates in bovine pancreatic trypsin inhibitor can be explained by two coexisting processes, in which the low activation-energy one does not correlate with the protein thermal stability in urea. The irrelevance of the hydrogen bonds in α-helix to the stability of this structure was argued by Lipscomb [see pp. 177−178 in the Discussion of Levitt 1981)]; while breaking a hydrogen bond in α-helix, NH and CO groups requires the formation of hydrogen bonds with a water molecule.

Correlations of $\sigma^2(r_0)$ with Other Structural Properties

The fact that most of the less-stabilized parts of the secondary structure are located near the surface of the proteins led to the prediction that σ^2 decreases towards the protein interior (Kuntz 1972). This correlation has been found by Sternberg et al. in lysozyme (1979). These authors predicted that σ^2 will increase as the square of the distance from the protein center of gravity. They tested the relation $\sigma^2(r_0) = A + Bd^2(r_0)$ where $d(r_0)$ is the distance between a C^α atom of residue r_0 and the molecular centroid and A and B are phenomenological parameters. The correlation coefficient was found to be 0.64. In our study relation $\sigma^2(r_0) = \sigma_0^2/\varrho(r_0)$ found to hold with correlation coefficient of 0.42 for lysozyme and 0.63 for metmyoglobin. However, in both cases the use of linear regression is not too meaningful if one consider as a positive result a pattern in $1/\varrho(r_0)$ or $d^2(r_0)$ that looks like a similar one in $\sigma^2(r_0)$ but is somewhat shifted, e.g., the situation in lysozyme around $r_0 = 45$ and 70 and in myoglobin, around $r_0 = 40$ and 75. These shifts contribute considerably to lowering the value of the correlation coefficient. In the work of Sternberg et al. (1979) the authors did not present $d^2(r_0)$ as a function of r_0 so one is unable to see if $d^2(r_0)$ prossesses the patterns of $\sigma^2(r_0)$. However, the authours denoted r_0 of the residues which are the most deviant from the predicted linear relation between $\sigma^2(r_0)$ and $d^2(r_0)$. These points happen to be in the peak regions of $\sigma^2(r_0)$. Therefore, an objective comparison between our results and those of Sternberg et al. (1979) is still

lacking. In contrast with the above low values of the correlation coefficient Eq. (3d) fits the data of the two proteins with a correlation coefficient of $> 0{,}99$. This presentation is insensitive to small shifts. We may conclude that it is likely that the shifts reflect a structural property but I have no explanation for their origin. It stands to reason to ask if the above prediction of Sternberg et al. and the present one contradict each other: The high $\sigma^2(r_0)$ regions are related by our theory to the less-stabilized regions in terms of lower density of stabilizing interactions which are located, as mentioned above, near the protein surface. Thus, the present theory is capable of explaining the decrease of σ^2 towards the protein interior as well. Both approaches are consistent with the calculated decrease of volume fluctuations towards the protein interior (Richards 1979).

Relevance of the Results to the Dynamics of Protein Structure

We already mentioned that $\sigma^2(r_0)$ is an equilibrium property, i.e., it reflects the magnitude of the protein structural fluctuations at r_0, averaged on time. The correlation of $\sigma^2(r_0)$ with the density of mainly hydrogen bonds rationalizes the following idea: *Some essential contributions to protein dynamics are likely to originate from the kinetics of formation and breaking of intramolecular hydrogen bonds and those resulting from protein-solvent interaction.* This idea is supported by few other studies: Ultrasound absorption studies in proteins solutions, probe structural rearrangement and/or volume fluctuations derived by chemical reactions, on time scale of $10^{-5}-10^{-9}$ s, which can be attributed to proton transfer processes (Schneider et al. 1969; O'Brien and Dunn 1972; Slutsky et al. 1977). Formation and breaking of hydrogen bonds is the mechanism of the proposed role of water molecules as "mobility catalyzers" of the protein structure (Chiragadze and Ovsepyan 1972). This type of kinetics has been demonstrated by computer simulations (Levitt 1981) on time scale of $10^{-11}-10^{-13}$ s. The increase of the oxygen-hydrogen distance from 1 Å to 3–4 Å upon breaking of a hydrogen bond [see Levitt (1981) in p. 173] is a good reasoning for "generating" structural fluctuations by kinetics of hydrogen bonding. The possible relevance of such correlated fluctuations to catalysis has already been suggested by Careri (1974).

Local flexibility of the polypeptide chain is expected to be expressed, in extreme cases, as 'local disorder'. This is the situation with the X-ray structure of antibodies (Colman et al. 1976), in which the segmental flexibility seems to be an important factor in the comformational transitions occurring as the result of hapten or antigen binding (Pecht 1982). The contribution of our work to these aspects of the protein function is a prediction that *some dynamic features of the antibody function should be linked to the kinetics of breaking and the formation of hydrogen bonds.* Another aspect of interest is the local flexibility, which is expressed by $\sigma^2(r_0)$, near the active site of lysozyme. Here, the substrate forms hydrogen bonds with the main chain at $r_0 = 59$ and 107 and with Trp 72, Trp 63, and Asp 101 (Phillips 1966). Residues Asp 52 and Glu 35 are involved in the bond cleavage (Chipman and Sharon 1969). Besides Asp 101, in all the other

locations $\sigma^2(r_0)$ exceeds small or moderate values. However, the fact that the X-ray analysis involved an inhibitor, which per definition is tightly bound, makes a correlation study somewhat ambiguous. One can at least conclude that the value of the presented analysis in terms of H-bonds and flexibility could be tested if the value of $\sigma^2(r_0)$ of lysozyme were available without the presence of an inhibitor.

The good correlation between $\sigma^2(r_0)$ and the density of stabilizing interactions, acting as constraints on a random chain, seems to justify the using of a simplified mechanistic approach for the description of the equilibrium and the dynamical properties of the main-chain structural fluctuations in proteins.

Appendix

Derivation of f(k) for a Random Chain

A given infinite random chain $(-\infty < r < \infty)$ is fully stabilized for the units $r \geq 0$, i.e.,

$$n(r) = \begin{cases} p & r \geq 0 \\ 0 & r < 0 \end{cases}. \tag{A-1}$$

Let $U_k(t)$ be the displacement of the k unit with respect to the $k-1$ one at time t. σ^2 of the l^{th} unit, relative to the zeroth one is

$$\sigma^2(-l) = \left\langle \left[\sum_{k=1}^{-l} U_k(t) \right]^2 \right\rangle_T = \sum_{k=1}^{-l} \sum_{k'=1}^{-l} \langle U_k(t)\, U_{k'}(t) \rangle_T ,$$

where the brackets denote a thermal averaging. Since different units are assumed to be independent and identical, $\langle U_k(t)\, U_{k'}(t) \rangle_T = U_0^2\, \delta_{kk'}$, where $U_0^2 = \langle U_k^2(t) \rangle$ *and* $\delta_{kk'}$, is 1 for $k = k'$ and 0 for $k \neq k'$. Thus

$$\sigma^2(-l) = U_0^2 l . \tag{A-2}$$

This result holds as well for a one-dimensional chain subject to harmonic forces (Imry and Gavish 1974) or entropic forces (de Gennes 1967) with or without the presence of viscous damping. Using Eqs. (3c) and (A-2) we find $\sigma^2(-l) = U_0^2 l = \sigma_0^2/\varrho(-l)$. Taking into account Eqs. (A-2) and (4) we find:

$$\sum_{k=0}^{\infty} f(k+l) = A/l \qquad \sum_{k=1}^{\infty} f(k+l) = A/(l+1) , \tag{A-3}$$

where $A = \sigma_0^2/U_0^2$. The right hand side expression was obtained from the left hand side one by the transformation $l \to l+1$. The difference between these two equations yields $f(l) = A/l(l+1)$. Taking into account the normalization condition, we find for the fully stabilized case, i.e., $n(k) = p$, that

$$1 = f(0) + 2 \sum_{k=1}^{\infty} f(k) = f(0) + 2A \sum_{k=1}^{\infty} 1/k(k+1) .$$

Since

$$\sum_{k=1}^{N} 1/k(k+1) = N/(N+1) .$$

Therefore $f(0) = 1 - 2A$ and

$$f(k) = \begin{cases} 1 - 2A & k = 0 \\ A/|k|(|k|+1) & k \neq 0 \end{cases} . \tag{A-4}$$

We took into these considerations implicity the fact that $f(r-r_0)$ is independent of r_0 and $n(r_0)$, and is a symmetric function of $r-r_0$.

It is of interest to calculate $\varrho(r_0)$ at the middle of an unstabilized segment of length l for which $l/2$ is an integral number (see Fig. 3B). Thus,

$$n(k) = \begin{cases} 0 & |k| < l/2 \\ 1 & |k| \geqslant l/2 \end{cases} \tag{A-5}$$

$$\varrho(0) = 2 \sum_{k=l/2}^{\infty} f(k) = 2A \left[\sum_{k=1}^{\infty} - \sum_{k=1}^{l/2-1} \right] f(k) = 2A[1 - (l/2-1)/(l/2)] = 4A/l .$$

Thus using Eq. 30 we find for this case

$$\sigma^2(0) = U_0^2 l/4 . \tag{A-6}$$

Acknowledgements. I thank Professors H. Frauenfelder, E. Gratton, B. Honig, F. M. Richards, J. L. Sussman, G. Weber and C. K. Woodwards for their helpful comments and criticism, and to Professor G. A. Petsko for supplying the metmyoglobin data. This work was supported by NSF grant PCM79-05072 and HEW-18051.

References

Artmyiuk P, Blake CCF, Grace DEP, Oatley SJ, Phillips DC, Sternberg MJE (1979) Crystallographic studies of the dynamic properties of lysozyme. Nature 280: 563–568

Careri G (1974) The fluctuating enzyme. In: Korsunoglu B, Mintz SL, Widmayer SM (eds) Quantum statistical mechanics in the natural sciences. Plenum Press, New York, pp 15–35

Chipman DM, Sharon N (1969) Mechanism of lysozyme action. Science 165: 454–465

Chiragadze YN, Ovsepyan AM (1972) Hydration mobility in peptide structures. Biopolymers 11: 2179–2186

Colman PM, Deisenhofer J, Huber R (1976) Structure of the human antibody molecule Kol (immuneglobulin Gl): An electron density map at 5 Å resolution. J Mol Biol 100: 257–282

de Gennes P-G (1967) Quasi-elastic scattering of neutrons by dilute polymer solutions: I. Free-draining limit. Physics 3: 37–45

Frauenfelder H, Petsko GA, Tsernoglou D (1979) Temperature-dependent X-ray diffraction as a probe of protein structural dynamics. Nature 280: 558–563

Gavish B (1978) The role of geometry and elastic strains in dynamic states of proteins. Biophys Struct Mech 4: 37–52

Gavish B (1981) Modelling the unusual temperature dependence of atomic displacements in proteins by local nonharmonic potentials. Proc Natl Acad Sci USA 78: 6868–6872

Hilton BD, Trudeau K, Woodward CK (1981) Hydrogen exchange rates in pancreatic trypsin inhibitor are not correlated to thermal stability in urea. Biochemistry 20: 4697–4703

Imry Y, Gavish B (1974) Correlation functions and structure factors for a lattice in a viscous medium. J Chem Phys 61: 1554–1558

Janin J, Wodak S, Levitt M, Maigret B (1978) Conformation of amino acid side-chains in proteins. J Mol Biol 125: 357–386

Kuntz ID (1972) Tertiary structure in carboxipeptidase. J Am Chem Soc 94: 8568–8572

Levitt M (1981) Hydrogen bond and internal solvent dynamics of BPTI protein. Ann NY Acad Sci 367: 162–181

McCammon JA, Karplus M (1980) Simulation of protein dynamics. Annu Rev Phys Chem 31: 29–45

O'Brien WD Jr, Dunn F (1972) Ultrasound absorption mechanisms in aqueous solutions of bovin hemoglobin. J Phys Chem 76: 528–533

Pecht I (1982) Dynamic aspects of antibody function. In: Selah M (ed) The antigen. Academic Press, New York, pp 1–68

Perutz MF (1978) Electrostatic effects in proteins. Science 201: 1187–1191

Phillips DC (1966) The three dimensional structure of an enzyme molecule. Sci Am (5) 215: 78–90

Richards FM (1979) Packing defects, cavities, volume fluctuations, and access to the interior of proteins. Including some general comments on surface area and protein structure. Carlsberg Res Commun 44: 47–63

Schneider F, Muller-Landay F, Mayer A (1969) Acoustical properties of aqueous solutions of oxygenated and deoxygenated hemoglobin. Biopolymers 8: 537–544

Slutsky LJ, Madsen L, White RD (1977) Acoustic absorption in solutions of small peptides and proteins. de Klerk J, McAvoy BR (eds) IEEE Ultrasonic Symposium (Proc.), pp 148–152

Stapleton HJ, Allen JP, Flynn CP, Stinson DG, Kurtz SR (1981) Fractal form of proteins. Phys Rev Lett 45: 1456–1459

Sternberg MJE, Grace DEP, Phillips DC (1979) Dynamic information from protein crystallography. J Mol Biol 130: 231–253

Takano T (1977) Structure of myoglobin refined at 2.0 Å resolution. J Mol Biol 110: 537–568

Zundel G (1976) Easily polarizable hydrogen bonds – their interaction with the environment – IR continoum and anomalous large proton conductivity. In: Schuster P, Zundel, G, Sandorfy C (eds) The hydrogen bond. North Holland, Amsterdam, Vol 2, pp 683–766