ELSEVIER

Biophysical Chemistry 53 (1994) 37–44

Biophysical Chemistry

The observation of structural transitions of a single protein molecule

K.F. Bradley [a,*], S.-H. Chen [a], M.-C. Bellissent-Funel [b], H.L. Crespi [c]

[a] *24-211 Massachusetts Institute of Technology, Cambridge, MA 02139, USA*
[b] *Laboratoire Léon Brillouin (CEA-CNRS), CE Saclay, 91191 Gif-sur-Yvette, Cedex, France*
[c] *Argonne National Laboratory, Argonne, IL 60439, USA*

Received 30 October 1993; accepted in revised form 30 December 1993

Abstract

Coherent neutron scattering measurements of an amorphous, in vivo deuterated *C-phycocyanin* are compared with a calculation of the individual protein molecule's coherent static structure factor. Both show the significant features associated with known structure factors of several amorphous materials, most notably, an unusually sharp first diffraction peak occurring near 1.4 $Å^{-1}$. We show that in the protein, such a peak results from the product of a form factor associated with correlations of atoms within individual amino acids and a structural term expressing inter-amino-acid correlations. The measurement, interpreted through behavior of the first diffraction peak, indicates that inter-amino-acid correlations – a measure of the protein's medium-range structure – undergo transitions which are primarily related to hydration rather than to temperature.

Keywords: *C-phycocyanin* protein; Neutron scattering; Structure factor; Transition

1. Introduction

Structural studies of proteins generally begin with a determination of the linear sequence of the constituent amino acids of the protein and then proceed to fix all the atomic coordinates in the protein molecule through refinement of X-ray diffraction data. This refinement relies on correlations between intermolecular pairs of atoms in a crystal of protein molecules. A particular protein of interest is *C-phycocyanin* which contains light absorption centers central to the type II photosynthesis performed by cyanobacteria (blue-green algae). As part of their effort to understand the photosynthetic process, Zuber and his collaborators have established the amino acid sequence of *C-phycocyanin* from several different species of cyanobacteria [1–4].

Based on this work, Schirmer and collaborators used their measured X-ray diffraction patterns of molecular crystals formed from *C-phycocyanin* to refine the coordinates of all the non-hydrogen atoms in the protein to a resolution of 2.1 Å [5–7].

The crystallized *C-phycocyanin* unit cell ($P6_3$, $a = b = 154.6$ Å, $c = 40.5$ Å) contains a protein hexamer of molecular weight 229346 Dalton. The more easily obtained powder form of the protein extracted directly from the cyanobacteria does not contain repeated unit cells, but rather is made up of individual hexamers with random orientations and a

* Corresponding author.

SSDI 0301-4622(94)00074-T

uniform density of about 0.3 g/ml. Such a sample is similar to amorphous solids and neutron diffraction data shows several broad diffraction peaks characteristic of noncrystalline solids. These broad diffraction peaks arise from correlations between pairs of atoms within the protein molecule.

As pointed out by Moss and Price [8], basic similarities in many glass forming structures give rise to quite similar diffraction patterns in a wide variety of glasses. In particular, the diffraction patterns of these glasses exhibit a low-q peak in the 1 to 1.4 $\AA^{-1}$ region which depends on temperature, pressure and glass composition. This dependence is quite different from the other diffraction peaks occurring above 2 $\AA^{-1}$. Often, upon cooling a glass, there is a moderate narrowing of the diffraction peaks due to a reduction of thermal vibrations. This low-q peak, however, can exhibit large changes in amplitude with temperature; Moss and Price associate these changes with interference between a form factor resulting from the internal structure of individual building blocks and the structure factor representing the spatial correlations between building blocks. For this reason the peak, called the first sharp diffraction peak or FSDP, is a sensitive measure of the medium range order of the three-dimensional random network. In proteins there is an analogous situation. The basic building blocks of proteins are the 20 different kinds of amino acid residues. We will show that, to a first approximation, the intra-residue form factors for the different residues can be expressed as a single average form factor. The inter-residue correlations vary with the secondary and tertiary structure of the protein, and the result analysed using the average form factor remains sensitive to these longer range structures in a protein molecule. In Section 3, we present the calculation of the structure factor of amorphous protein using the glass model [8] and compare it with the experimental structure factor deduced from neutron scattering experiments [9] described in Section 4.

2. Materials and methods

2.1. Sample preparation

The *C-phycocyanin* protein is a light-harvesting protein abundant in blue-green algae. Nearly 99% deuterated samples of this phycobiliprotein were isolated from the cyanobacteria *Synechococcus lividus* which were grown in perdeuterated cultures [10] (99% pure D_2O) at Argonne National Laboratory, Argonne, IL.

This process yielded deuterated protein that had virtually all its ^1H–C bonds replaced by ^2H–C bonds. Deuterium in the weaker H–N and H–O bonds will tend to exchange with atmospheric hydrogen after extraction; however, we minimized the amount of hydrogen in these bonds by dissolving the protein in D_2O and freeze-drying it in a D_2O rich atmosphere. We then either sealed the samples in the standard 7C2 sample containers or adsorbed the appropriate amount of D_2O onto the protein and then sealed it into the sample containers. At full hydration the protein *C-phycocyanin* contains 0.5 gram of D_2O per gram of protein. The water content was measured by the increase in weight of the protein sample after exposing the protein to the vapor of D_2O. We prepared two samples with different degrees of hydration. The first one was 73% hydrated, i.e. contained 0.365 gram D_2O per gram of protein and the second one was 35% hydrated, i.e. it contained 0.175 gram of D_2O per gram of protein. These samples contain these amount of water in addition to the 4% of water molecules (D_2O), which have to be considered like in many other proteins, as an integral part of the molecule.

2.2. Neutron scattering experiments

In order to explore the dependence of the protein static structure factor on temperature and hydration, we performed a series of diffraction experiments using the 7C2 spectrometer, at the Orphee reactor of the Laboratoire Leon Brillouin, Saclay, France. The 7C2 spectrometer is equipped with a BF3 position sensitive detector with 640 cells; the angular step between two adjacent cells is equal to 0.2 deg which leads to a maximum diffraction angle of 128 deg. We selected an incident wavelength of $\lambda = 0.703$ Å by means of a Cu (111) monochromator that allowed us to cover a range of magnitude of scattering wavevectors ($Q = 4\ \pi\ \sin\ \theta/\lambda$) extending from 0.7 to 16 $\AA^{-1}$. The protein samples were held in containers of thin-walled vanadium (0.1 mm thick) with an internal diameter equal to 6 mm and placed inside

a cryostat equipped with a vanadium tail. The experiments were done at several temperatures between 295 and 77 K. For the 'dry' *C-phycocyanin* samples and the two hydrated *C-phycocyanin* (35% and 73%) samples, the following sequence of measurements: 'sample + container' at each required temperature, 'empty container', 'standard vanadium bar', 'empty cryostat', and 'cadmium rod' having the same dimensions as the samples were made. After the usual attenuation and multiple scattering corrections and the normalisation procedure using the standard vanadium bar, we obtained the total differential scattering cross section. After correcting for the inelasticity effect [9] the structure factor $S(q)$ of the protein sample was evaluated. Fig 1 gives an example of two $S(q)$ plots for the protein *C-phycocyanin* at 295 K for two levels of hydration. Comparison is made with the dry protein sample.

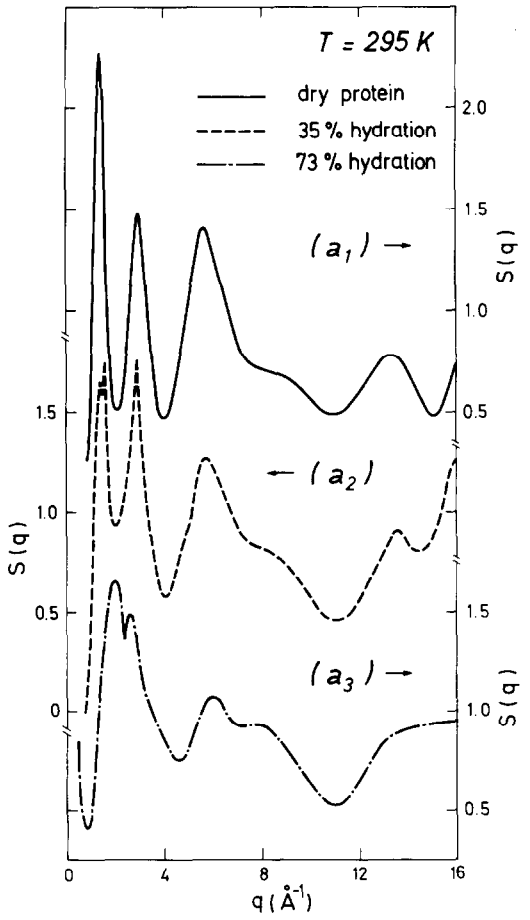

Fig. 1. $S(q)$ plots for a deuterated *C-phycocyanin* protein at $T = 295$ K for two different levels of hydration (35% (a_2) and 73% (a_3)). Comparison is made with the dry sample (a_1). It is seen that the primary effects of hydration are to alter the peak structures in the q-range below 3 Å^{-1}.

3. The structure factor $S(q)$ of amorphous protein

3.1. Calculation of $S(q)$ from X-ray scattering atomic coordinates

The static coherent differential cross section measured in a diffraction experiment is given by:

$$\frac{d\sigma_c}{d\Omega} = N_0\langle b^2\rangle S(q). \quad (1)$$

$\langle b^2\rangle$ is the average of the squared-neutron-scattering length of the scattering object which in this case, is the individual protein molecule. N_0 is the number of scattering objects in the sample. $S(q)$ is the q-space correlation function for the scattering object which is a three-dimensional Fourier transform of the atom–atom pair correlation function.

Using the atomic coordinates supplied by Duerring and Huber from their X-ray work, we have calculated the static structure factor of an individual molecule of *C-phycocyanin* according to its definition (Eq. (2)). The dotted line in Figs. 2a and 2b is the calculated $S(q)$ which exhibits all the general characteristics of measured structure factors of amorphous solids. At large q, the curve goes to $\langle b\rangle^2/\langle b^2\rangle$. The broad peaks at intermediate q range, down to about 2.0 Å^{-1}, arise out of correlations due to favorable nearest neighbor bond lengths and other local structures.

3.2. Calculation of $S(q)$ from models of glass systems

In order to evaluate the structural data in terms of the static amino acid structures within the protein, we can extend the ideas of Price (unpublished results, 1983), developed for glass systems with a single structural forming unit, to attain an expression for $S(q)$ which can be separated into two terms; one describing the structure of the individual amino acid residues and one describing the correlations between

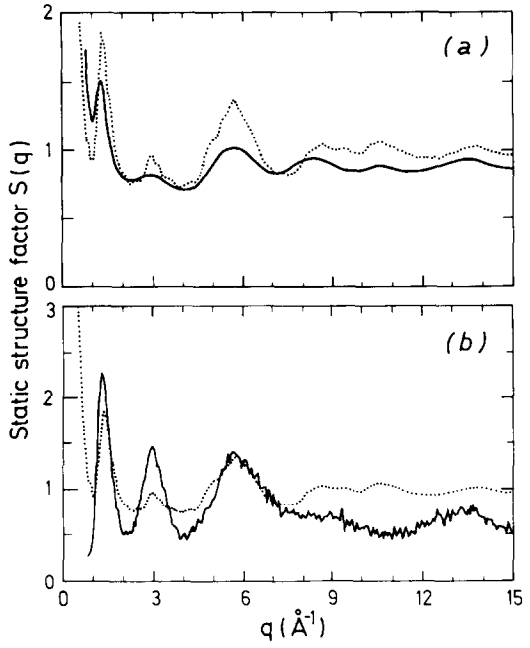

Fig. 2. (a) Static structure factors calculated from the same atomic coordinate data of a *C-phycocyanin* molecule. The dotted line is the result found using the definition of $S(q)$ and the solid line is calculated using Eqs. (8), (7), (3), and (9) in the text. (b) Comparison of the measured $S(q)$ of dry protein *C-phycocyanin* from *Synechococcus lividus* (solid line) and the result of calculating $S(q)$ using the atomic coordinates of *C-phycocyanin* (dashed line).

these residues. The function $S(q)$ in Eq. (1) can be expressed using the definition of the pair correlation function to include each atom's position $\boldsymbol{u}_i$. By taking the equilibrium position of the C_α atom in the kth amino acid as $\boldsymbol{R}_k$, each atomic position in that residue can be written as $\boldsymbol{u}_{i\in k} = \boldsymbol{R}_k + \boldsymbol{r}_i$ and the protein static structure factor is written as

$$S(q) = F_1(q) + \frac{1}{N\langle b^2\rangle} \sum_{\substack{k \\ l\neq k}} \langle e^{iq\cdot(\boldsymbol{R}_k - \boldsymbol{R}_l)}\rangle \times \sum_{\substack{i\in k \\ j\in l}} \langle b_i\rangle\langle b_j\rangle\langle e^{iq\cdot(\boldsymbol{r}_i - \boldsymbol{r}_j)}\rangle. \quad (2)$$

b_i is the neutron scattering length of the target nuclei found at position $\boldsymbol{u}_i$, N is the number of atoms in the protein, N_r is the number of residues or amino acids in the protein molecule, k and l index amino acids, and i and j index the atoms within an amino acid. $F_1(q)$ is the average residue form factor which is associated with the positions of pairs of atoms within an 'average' residue:

$$F_1(q) = \frac{N_r}{N\langle b^2\rangle} \left\langle \sum_{i,j\in k} b_i b_j \, e^{iq\cdot(\boldsymbol{r}_i - \boldsymbol{r}_j)} \right\rangle. \quad (3)$$

The second term in Eq. (2) includes a 'residue–residue' structure factor associated with the positions of the various residues to one another. If $\boldsymbol{r}_i$ is independent of $\boldsymbol{R}_k$ then the two sums in the second term of Eq. (2) could be separated yielding an $S(q)$ which explicitly depends on a residue–residue structure factor. However, unlike glasses formed from identical building blocks, the amino acids of any protein are not all identical and so $\boldsymbol{r}_i$ does indeed depend on $\boldsymbol{R}_k$.

This dependence can be broken by introducing another set of indices (m and n) to indicate each of the 20 residue types. With this additional index, $S(q)$ is now formed by three terms:

$$S(q) = F_1(q) + \sum_{m,n\neq m} F_2^{m,n}(q) S_{r-r}^{m,n}(q) + \sum_m F_2^{m,m}(q)\left[S_{r-r}^{m,m}(q) - 1\right]. \quad (4)$$

$S_{r-r}^{m,n}(q)$ is the residue–residue structure factor associated with the positions of residue pairs made up of type m and type n residues,

$$S_{r-r}^{m,n}(q) = \left\langle \frac{1}{N_r} \sum_{\substack{k\in m \\ l\in n, l\neq k}} e^{iq\cdot[\boldsymbol{R}_k - \boldsymbol{R}_l]} \right\rangle. \quad (5)$$

$F_2^{m,n}(q)$ is a weighting factor associated with the positions of pairs of atoms which are members of *different* residues m and n,

$$F_2^{m,n}(q) = \frac{N_r}{N\langle b^2\rangle} \left\langle \sum_{\substack{i\in m \\ j\in n, j\notin R(i)}} b_i b_j \, e^{iq\cdot[\boldsymbol{r}_i - \boldsymbol{r}_j]} \right\rangle, \quad (6)$$

where $R(i)$ is the residue in which atom i resides.

Having established Eq. (4) we can ask if the rigor maintained is worth the computational complexity of an equation with 210 partial structure factors and their associated 210 weighting factors. In fact, while each of the 20 different residues do have a different structure, they also share many structural characteristics. For instance, all the residues have identical

backbone structures which, for several of the residues, comprise the majority of the their atoms. We will now explore the degradation of the model imposed by the assumption that all the residues have identical structures.

In this assumption there is only one residue type $m = n = 1$ leading to

$$F_2^{m,n\neq m}(q) = 0,$$

$$F_2^{m,m}(q) = F_2(q) = \frac{N_r}{N\langle b^2\rangle}\left\langle \sum_{\substack{i \\ j\notin R(i)}} b_i b_j \, e^{iq\cdot[r_i - r_j]}\right\rangle,$$

$$S_{r-r}^{n,m}(q) = S_{r-r}(q) = \left\langle \frac{1}{N_r}\sum_{\substack{k \\ l\neq k}} e^{iq\cdot[R_k - R_j]}\right\rangle, \qquad (7)$$

and Eq. (4) reduces to

$$S(q) = F_1(q) + F_2(q)\left[S(q)_{r-r} - 1\right]. \qquad (8)$$

Price (unpublished results, 1983) points out that for randomly oriented building blocks, equations in the form of $F_2(q)$ averages to

$$F_2(q) = \frac{N_r}{N\langle b^2\rangle}\left\langle \sum_i b_i \, e^{iq\cdot r_i}\right\rangle^2. \qquad (9)$$

4. Results and discussion

4.1. Comparison between S(q) derived from X-ray scattering atomic coordinates and S(q) from the glass model

In Fig. 2(a) we show $S(q)$ (dotted line) for the individual protein molecule calculated using direct atomic positions in one case and Eqs. (8), (7), (3), and (9) in the other case (solid line). This comparison makes clear that the use of Eq. (8) captures the essential features of the $S(q)$ of the molecule. But the neglect of the fine structure allowed by the individual partial residue structure factors degrades the results of the calculation. In spite of this degradation, the curve calculated using (8) contains the important distinction between the sharp low-q peak and the other, higher-q, peaks. It is the higher-q peaks which are most degraded by the uniform residue assumption. This is because the sharp features in the peaks above 2 Å^{-1} are all due to irregularities introduced by the variable intra-residue correlations and these have been smoothed through the use of Eq. (8).

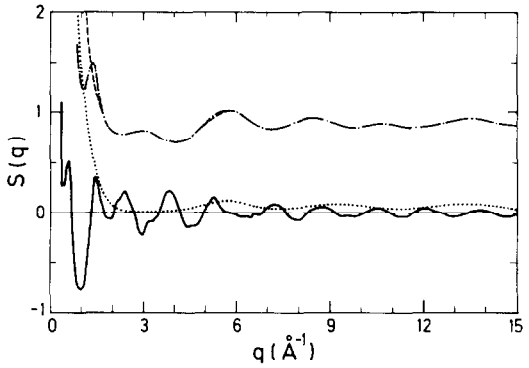

Fig. 3. The functions $S(q) = F_1(q) + F_2(q)[S_{r-r}(q) - 1]$ (chain-dot line, upper figure), $S_{r-r}(q) - 1$ (solid line, lower figure), $F_1(q)$ (dotted line, upper figure), and $F_2(q)$ (dashed line, upper figure) as defined in the text. Notice in the figure for $S_{r-r}(q) - 1$, there are two major peaks in the q-range below 3 Å^{-1}, one at 1.34 Å^{-1}, the other at 2.2 Å^{-1} (see lower figure).

The source of this smoothing is clear when the constituent functions in Eq. (8) are calculated separately. Fig. 3 shows the functions $F_1(q)$, $F_2(q)$, $[S_{r-r}(q) - 1]$ and $S(q) = F_1(q) + F_2(q)[S_{r-r}(q) - 1]$. This indicates that the source of the peak at 1.34 Å^{-1} results from the product of the rapidly changing $F_2(q)$ and oscillatory $[S_{r-r}(q) - 1]$ functions near this q-value. It is only below about 2 Å^{-1} that $F_2(q)$ is large enough to amplify the $[S_{r-r}(q)$-1] function so that it noticeably contributes to $S(q)$. In addition, it is only above about 0.5 Å^{-1} that $[S_{r-r}(q) - 1]$ contains information on the medium-range order of amorphous protein. (Below 0.5 Å^{-1} $[S_{r-r}(q) - 1]$ is dominated by long-range order effects which, for proteins, are controlled, in particular, by the overall molecular shape.)

4.2. Comparison between S(q) from X-ray scattering atomic coordinates and S(q) from neutron scattering experiments

Fig. 2b is a comparison between the measured $S(q)$ of the dry protein (solid line) and the full $S(q)$ calculated directly from the atomic positions of the protein according to Eq. (2) (dotted line). There are

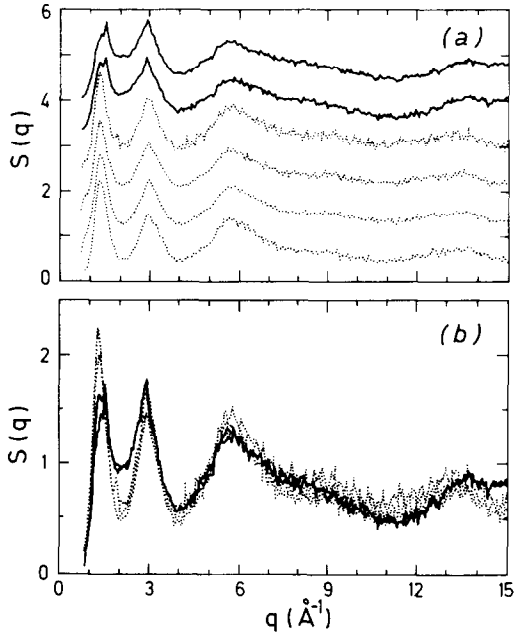

Fig. 4. (a) Measured $S(q)$ of dry *C-phycocyanin* at $T = 295$, 200, 77, and 20 K protein (dotted lines in order of increasing offset each of 0.8 structure-factor units); measured $S(q)$ of *C-phycocyanin* hydrated with 35% by weight D_2O at 295 and 77 K (solid lines in order of increasing offset each of additional 0.8 structure-factor units). (b) Same plot as (a), but without offsetting the curves.

significant differences between the two curves. Most of the disagreement arises from the fact that the coordinates determined with X-ray scattering have a resolution of 2.1 Å, an uncertainty which can introduce significant error into the calculated single-protein $S(q)$. Nonetheless, Fig. 2b indicates that the basic features of the measurement are captured in the calculation based on the atomic coordinates measured in a crystalline protein.

4.3. Interpretation of S(q) from neutron scattering experiments in light of the glass model

Fig. 4a shows four structure factors of the dry protein at temperatures 20, 77, 200 and 295 K as well as two $S(q)$ of protein hydrated at 35% with D_2O. The hydrated protein was measured at 295 and 77 K. In order to better see the differences between these spectra, all six curves are plotted in Fig. 4b without any vertical offset. This figure shows clearly that the only differences in $S(q)$ across this temperature and hydration range, aside from noise, occur below 3 Å^{-1}. In fact, temperature changes predominantly affect $S(q)$ of the dry protein only by changing the intensity of the 1.34 Å^{-1} peak in the structure factor. Comparison with Fig. 3 shows that, while the 1.34 Å^{-1} peak primarily results from a rising value of $[S_{r-r}(q) - 1]$ with decreasing q, the peaks at 2.96 and 5.65 Å^{-1} are due to intra-residue correlations ($F_1(q)$). For this reason, the 2.96 and 5.65 Å^{-1} peaks may change their intensity with temperature, but should not alter their position or substructure because the individual residue structures are not varying.

The solid lines in Fig. 4b, which are the $S(q)$ for the hydrated protein, show that the hydrated protein has undergone a significant structural transition with hydration. While the 1.34 Å^{-1} peak has split into two separate peaks, there are other changes that can best be seen when expressly viewed as a difference from the dry $S(q)$. Using $S(q)$ of the dry protein at $T = 295$ K as a standard, we plot the difference between this standard and the other measured values of $S(q)$ in Fig. 5. This plot confirms that the temperature-dependent effects in the dry protein are limited to changes in the 1.34 Å^{-1} peak intensity and that the hydrated protein, at both room temperature and 20 K, has an $S(q)$ significantly less intense at this momentum transfer than any of the dry protein structure factors. More importantly, Fig. 5 shows that the

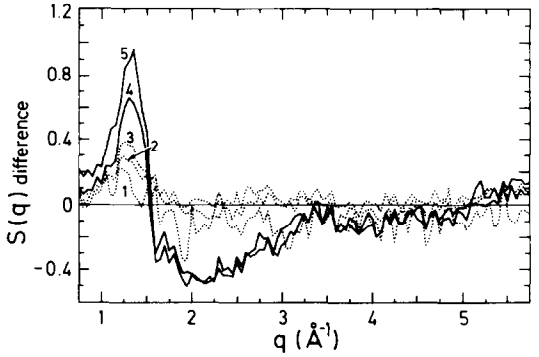

Fig. 5. The difference between the $S(q)$ of dry *C-phycocyanin* at $T = 295$ K and the dry protein $S(q)$ at 200, 77 and 20 K (dotted lines labeled 1, 2 and 3 respectively) and the difference between the same standard and the hydrated protein $S(q)$ at 295 and 77 K (solid lines labeled 4 and 5 respectively).

dominant effect of hydrating the protein is a large increase in intensity centered at 2.2 $\mathring{A}^{-1}$ and ranging from 1.6 to 3 $\mathring{A}^{-1}$. This region is below the peak centered at 3 $\mathring{A}^{-1}$ visible on both the measured and calculated spectra (see Fig. 2b). In addition, it cannot result from a shift in $F_2(q)$ to higher q-values because such a shift would increase the intensity of $S(q)$ at the 1.34 $\mathring{A}^{-1}$ peak as well as in the 2.2 $\mathring{A}^{-1}$ region. (Such a shift in $F_2(q)$ would be equivalent to a rearrangement of the average orientation of the residues to one-another.) This evidence indicates that the effect of hydrating the protein is primarily to increase the 2.2 $\mathring{A}^{-1}$ peak intensity in the $S_{r-r}(q)$ spectrum (see Fig. 3). This peak, which is less intense, but significantly broader, than the 1.34 $\mathring{A}^{-1}$ peak is therefore seen to represent correlations of important significance for the biological activity of this protein.

5. Conclusion

This work demonstrates that by interpreting the structure of an individual protein molecule as consisting of an amorphous arrangement of amino-acid residues, the effects of temperature and hydration on the structural modification of the protein can be characterized as changes in relative amino-acid positions. These changes can be then observed and quantified from the coherent scattering structure factor measured from an amorphous powder protein.

Acknowledgement

The authors would like to thank Marcus Duerring for kindly transmitting the *C-phycocyanin* atomic coordinates to us before processing could be completed by the Brook-haven Protein Data Bank. In addition, they take pleasure in acknowledging helpful conversations with P. Thiyagarajan and David Price. Financial support for this work was supplied by the United States Department of Energy.

References

[1] G. Frank, W. Sidley, H. Widmer and H. Zuber, Hoppe-Seyler's Z. Physiol. Chem. 359 (1978) 1491.
[2] W. Sidler, Jurg Gysi, E. Isker and H. Zuber, Hoppe-Seyler's Z. Physiol. Chem. 362 (1981) 611.
[3] P. Fuglistaller, F. Suter and H. Zuber, Hoppe-Seyler's Z. Physiol. Chem. 364 (1983) 691.
[4] R. Ruembeli, M. Wirth, F. Suter and H. Zuber, Biol. Chem. Hoppe-Seyler 368 (1987) 1.
[5] T. Schirmer, W. Bode, R. Huber, W. Sidler and H. Zuber, J. Mol. Biol. 184 (1985) 257.
[6] T. Schirmer, W. Bode and R. Huber, J. Mol. Biol. 196 (1987) 677.
[7] M. Duerring, R. Huber and W. Bode, FEBS Letters 236 (1988) 167.
[8] S.C. Moss and D.L. Price, Physics of disordered systems, eds. D. Adler, H. Fritzsche and S.R. Ovshinsky (Plenum Press, New York, 1985).
[9] M.-C. Bellissent-Funel, J. Lal, K.F. Bradley and S.H. Chen, Biophys. J. 64 (1993) 1542.
[10] H.L. Crespi, in: Stable isotopes in the life sciences (IAEA, Vienna, 1977) p. 111.