Structural hierarchy of several proteins observed by wide-angle solution scattering

Mitsuhiro Hirai,[a]* Hiroki Iwase,[a] Tomohiro Hayakawa,[a] Keiko Miura[b] and Katsuaki Inoue[b]

[a]Department of Physics, Gunma University, Maebashi 371-8510, Japan, and [b]Japan Synchrotron Radiation Research Institute, Sayo, Hyogo 679-5198, Japan.
E-mail: mhirai@fs.aramaki.gunma-u.ac.jp

In the present study using a high-intensity X-ray beam from a third-generation synchrotron radiation source, it is demonstrated that a wide-angle X-ray scattering (WAXS) profile from several globular proteins in solution can reflect not only the overall structures (~300 Å distance resolution) but also intramolecular structures ranging to secondary structures (~2.5 Å distance resolution). The proteins treated in the present experiments are classified as different types of structure categories, namely, as all-α, all-β and $\alpha + \beta$ proteins. Here the full-range experimental scattering curves are compared with the theoretical curves, suggesting a further availability of the SR-WAXS method for studies of structure hierarchy and the function of proteins in solutions.

Keywords: solution X-ray scattering; structural hierarchy; secondary structures; proteins.

1. Introduction

Proteins are well known to take hierarchical structures in the native state, which is deeply involved in the appearance of various biological functions of proteins. Solution X-ray scattering methods of proteins are exclusively used to determine these structures, such as rough shape, radii of gyration or some large-scale intramolecular heterogeneity, mainly by using data sets in small-angle scattering regions (Pilz, 1982; Feigin & Svergun, 1987). As solution scattering methods using X-rays and neutrons are well recognized to provide insights into biomolecular structure and function, the solution scattering techniques have been developed continuously and have served advanced applications in structural molecular biology (Trewhella *et al.*, 1998). In many cases of these applications, solution scattering experiments of proteins are carried out at a low protein concentration to avoid modulation of the small-angle scattering data regions caused by an aggregation or a repulsive intermolecular interaction. Because of this, experimental data observed in the high-angle scattering region have too low statistics to be used for analyses of local structures such as secondary structures.

Recently, by using small- and medium-angle synchrotron radiation X-ray scattering data, we successfully revealed a thermal structural transition multiplicity of a globular protein strongly dependent on the protein structure hierarchy (Hirai *et al.*, 1998, 2000; Hirai, Arai & Iwase, 1999; Arai & Hirai, 1999), indicating that the solution X-ray scattering method still serves us further fruitful information on protein structures. In the previous reports, we showed an availability of small- and medium-angle synchrotron radiation X-ray scattering to clarify a mechanism of folding-and-unfolding of proteins, where we presented a new method of the solution scattering data analysis combined with thermal analysis of differential scanning calorimetry (Hirai, Arai & Iwase, 1999; Hirai *et al.*, 2000). In the present study, to elucidate a further availability of the wide-angle synchrotron radiation X-ray scattering (SR-WAXS) method for solutions, we have carried out medium- and high-angle solution scattering experiments of several globular proteins which are classified as different structural types of proteins. Here we demonstrate that SR-WAXS data covering the q-range from ~0.02 Å^{-1} to ~2.5 Å^{-1} (where q is the absolute value of the scattering vector, $q = 4\pi \sin\theta/\lambda$, 2θ is the scattering angle and λ is the wavelength) are available to clarify not only an intramolecular domain correlation but also secondary structures of proteins.

2. Materials and methods

2.1. Samples

The proteins used for the measurements were myoglobin from horse skeletal muscle, hemoglobin from bovine, α-chymotrypsin from bovine pancreas, lysozyme from chicken egg-white, ribonuclease A from bovine pancreas and α-lactalbumin from bovine milk, which were all purchased from Sigma Chemical Co. These proteins belong to different types of protein structure categories, namely, to all-α, all-β or $\alpha + \beta$ proteins. To attain high statistical scattering data avoiding aggregation, we dissolved proteins in low-ionic-strength solvents. Here we used 10 mM HCl solvent for myoglobin, hemoglobin, lysozyme and α-chymotrypsin, and 50 mM HEPES [N-(2-hydroxymethyl)piperazine-N'-(2-ethanesulfonic acid)] solvent for α-lactalbumin and ribonuclease A. The final pH and concentration of the protein solutions were 5% w/v myoglobin at pH 6.4, 5% w/v hemoglobin at pH 5.0, 5% w/v α-lactalbumin at pH 7.0, 5% w/v lysozyme at pH 5.0, 5% w/v ribonuclease A at pH 7.0 and 1% w/v α-chymotrypsin at pH 4.

2.2. X-ray scattering measurements

Solution X-ray scattering experiments were carried out by using an X-ray scattering spectrometer installed at BL40 of the 8 GeV synchrotron radiation source at the Japan Synchrotron Radiation Research Institute (JASRI), Harima, Japan. The sample-to-detector distance was 46 cm. The X-ray wavelength used was selected to be 0.729 Å for the high-angle scattering measurements and 1.550 Å for the small-angle and medium-angle scattering measurements. The X-ray scattering intensity was detected by an imaging plate (IP) system from Rigaku R-AXIS IV using a 30 cm × 30 cm area of the IP. Incident and transmitted X-ray beam intensities were monitored by a pair of ionization chambers. Details of the spectrometer have been described elsewhere (Miura *et al.*, 2000). The sample solutions were contained in the sample cell with a 1 mm path length and a pair of thin-quartz windows. The exposure time to X-rays was 60 s. Two-dimensional scattering data recorded on the IP were transformed into a one-dimensional scattering intensity profile $I(q)$ by using a circle-averaging program.

3. Results and discussion

3.1. Wide-angle scattering curves of different types of proteins

In solution scattering measurements, the background correction of the scattering intensity $I(q)$ is conventionally made by using the following equation (Hirai *et al.*, 1993),

$$I(q) = \frac{1}{B_{\rm sol} T_{\rm sol}} I_{\rm sol}(q) - \frac{1}{B_{\rm solv} T_{\rm solv}} I_{\rm solv}(q), \qquad (1)$$

where $I_{\rm sol}(q)$, $I_{\rm solv}(q)$, $B_{\rm sol}$, $B_{\rm solv}$, $T_{\rm sol}$ and $T_{\rm solv}$ are the observed scattering intensities, the incident beam intensities, and the transmissions of the solution and solvent samples, respectively. In many cases for small-angle and medium-angle scattering data, the background

correction using equation (1) is used to obtain a net scattering intensity profile from the solute particles, whereas in the present case of high-angle scattering data above $q \simeq 1$ Å^{-1} we observed a strong peak at ~ 2 Å^{-1} originating from the correlation among water molecules. The background correction above $q \simeq 0.8$ Å^{-1} using equation (1) did not subtract the background scattering from the water exactly. As shown previously (Bosio *et al.*, 1989), the structure factor of water begins to increase above $q \simeq 0.8$ Å^{-1}, suggesting that equation (1) is not appropriate for the scattering data correction above $q \simeq 0.8$ Å^{-1} since equation (1) does not consider the excluded volume of the solute macromolecules. In other words, above $q \simeq 0.8$ Å^{-1} a water solvent should not be regarded as giving homogeneous background scattering, as discussed again in the following section. Then, above $q \simeq 0.8$ Å^{-1} we used the following equation which takes into account both the excluded volume effect of the protein and the normalization by the correlation peak intensity of water,

$$I(q) = \frac{1}{I_{\rm peak}^{\rm sol}}\left[\frac{1}{B_{\rm sol}T_{\rm sol}}I_{\rm sol}(q) - \frac{1}{B_{\rm cell}T_{\rm cell}}I_{\rm cell}(q)\right] - \frac{(1-cv_a)}{I_{\rm peak}^{\rm solv}}\left[\frac{1}{B_{\rm solv}T_{\rm solv}}I_{\rm solv}(q) - \frac{1}{B_{\rm cell}T_{\rm cell}}I_{\rm cell}(q)\right], \quad (2)$$

where $I_{\rm peak}^{\rm sol}$ and $I_{\rm peak}^{\rm solv}$ are the water-correlation peak intensities of the solution and the solvent, c and v_a are the concentration of the protein molecule and its partial specific volume, and $I_{\rm cell}(q)$, $B_{\rm cell}$ and $T_{\rm cell}$ are the observed scattering intensity, the incident beam intensity and the transmission of the sample cell, respectively. Here the v_a values used were 0.74 ml g^{-1} for myoglobin, 0.75 ml g^{-1} for hemoglobin, 0.74 ml g^{-1} for α-chymotrypsin, 0.72 ml g^{-1} for lysozyme and α-lactalbumin, and 0.71 ml g^{-1} for ribonuclease A. By using equation (2), we successfully corrected the effect of the correlation peak of water on the scattering data.

Fig. 1 shows the corrected scattering curves of the different types of proteins, where the high-angle scattering curves are connected to the small-angle and medium-angle scattering curves around $q \simeq 0.7$ Å^{-1}. To attain high statistic scattering data under unaggregative solvent conditions, we used protein solutions with relatively high concentrations. In some samples the observed scattering curves in the small-angle region below $q \simeq 0.08$ Å^{-1} show a slight decreasing tendency or a broad peak owing to the intermolecular correlation under repulsive interparticle interaction. Except for a concentrated system composed of extremely anisotropic-shaped particles forming liquid-crystal ordering (Hirai *et al.*, 1997), for globular particle systems in solutions the repulsive interaction among the particles usually produces a broad single peak in the small-q region, for which the width and height of the broad peak depend on the effective charge of the solute particle, which scarcely affects the scattering curve in the q region above the peak position (Hirai, Takizawa *et al.*, 1996; Hirai, Iwase & Hayakawa, 1999). This is the case for one of the present protein solutions, as has already been shown (Arai & Hirai, 1999). According to the Structure Classification of Proteins database (SCOP; Murzin *et al.*, 1995), myoglobin and hemoglobin are classified as all-α proteins; α-chymotrypsin as an all-β protein; and lysozyme, ribonuclease A and α-lactalbumin as $\alpha + \beta$ proteins. Table 1 summarizes the contents of the secondary structures of the proteins used for the present experiments and the Protein Data Bank (PDB) file codes used for the following theoretical calculations of the solution scattering curves of the proteins. The PDB files used are 1WLA from Maurus *et al.* (1997), 1HDA from Perutz *et al.* (1993), 4CHA from Tsukada & Blow (1985), 7RSA from Wlodawer *et al.* (1988), 1ALC from Acharya *et al.* (1989) and 6LYZ from Diamond (1974). In Fig. 1 we can recognize the differences in the scattering curves of those proteins in the q range from ~ 0.2 Å^{-1} to ~ 2.5 Å^{-1}, suggesting that the experimental medium- and high-angle scattering curves reflect the characteristics of the intramolecular structures of the proteins. Namely, the scattering curves in the q range from ~ 0.2 Å^{-1} to ~ 0.7 Å^{-1} and in the q range from ~ 1 Å^{-1} to ~ 2.5 Å^{-1} mostly correspond to the intramolecular domain structures and to the secondary structures (helix or sheet structure periodicity and their packing), respectively. Fig. 2 schematically shows this situation. In Fig. 2 the regions in the theoretical scattering curve of hen egg-white lysozyme (6LYZ) are roughly assigned to the typical intramolecular distances depending on the structural hierarchy. Here we calculate the theoretical scattering curve using the following Debye equation (Debye, 1915) and the atomic coordinates of the proteins given by the PDB,

Table 1
Summary of the contents of secondary structures of the proteins used for the present experiments and the PDB file codes used for the theoretical scattering curve calculations.

Protein	PDB file code	Helix content (%)	Sheet content (%)
Myoglobin	1WLA	73.9	0
Hemoglobin	1HDA	68.1	0
α-Chymotrypsin	4CHA	8.3	30.7
Ribonuclease A	7RSA	17.7	31.5
α-Lactalbumin	1ALC	30.1	6.5
Lysozyme	6LYZ	30.2	6.2

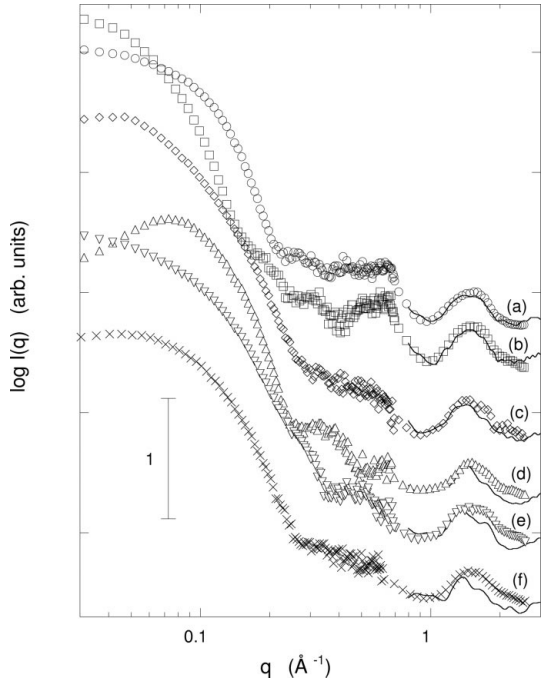

Figure 1
Wide-angle scattering curves of several globular proteins in solutions. (*a*) 5% *w*/*v* myoglobin at pH 6.4; (*b*) 5% *w*/*v* hemoglobin at pH 5; (*c*) 1% *w*/*v* α-chymotrypsin at pH 4; (*d*) 5% *w*/*v* lysozyme at pH 5; (*e*) 5% *w*/*v* ribonuclease A at pH 7; (*f*) 5% *w*/*v* α-lactalbumin at pH 7. These proteins are classified as different types of protein structure categories, namely, myoglobin and hemoglobin as all-α proteins, α-chymotrypsin as an all-β protein, and lysozyme, ribonuclease A and α-lactalbumin as $\alpha + \beta$ proteins. Full lines above $q > 0.8$ Å^{-1} show the theoretical solution scattering curves obtained by the Debye equation using the atomic coordinates of the proteins from the PDB.

$$I(q) = \sum_{i=1}^{n}\sum_{j=1}^{n} f_i f_j \frac{\sin(qr_{ij})}{qr_{ij}}, \qquad (3)$$

where f_i and f_j are the atomic scattering factors of the ith and jth atoms in a molecule composed of n atoms, and r_{ij} is the interatomic distance between the ith and jth atoms. The three-dimensional picture in Fig. 2 was drawn using the program *MolScript* (Kraulis, 1991). In Fig. 1 the theoretical scattering curves of the proteins from the Debye equation are superimposed onto the experimental curves in the q range from 0.8 Å^{-1} to 3 Å^{-1}. As the theoretical scattering curves of the proteins obtained by the Debye equation do not take into consideration a hydration shell surrounding the protein, the theoretical scattering curves in the q range below ~0.7 Å^{-1} greatly deviate from the experimental curve. However, the theoretical scattering curves in the q range from 0.8 Å^{-1} to 3 Å^{-1} reflect well the characteristics of the experimental curves depending on the secondary structures.

3.2. Comparison of experimental and theoretical scattering curves in the large-q range

To compare the experimental scattering curves over the large-q range with the theoretical curves we used the Debye equation for the high-angle scattering region ($q \gtrsim 0.8$ Å^{-1}) and the multipole expansion method (Stuhrmann & Miller, 1978) for the small- and medium-angle scattering region ($q \lesssim 0.7$ Å^{-1}). The program *CRYSOL*, which uses the multipole expansion method, is known to explain very well the experimental scattering curves of proteins in solutions, especially for the q region below ~0.5 Å^{-1}, by considering the excess average scattering density (so-called 'contrast') of the hydration shell (Svergun *et al.*, 1995, 1998). The concept of 'contrast' stands on the physical basis that solvent molecules are regarded as giving homogeneous background scattering; therefore this concept no longer holds in the high-angle scattering limit due to the finite volumes of solvent molecules. As mentioned by the developers, the *CRYSOL* program does not ensure correct theoretical solution scattering curves above ~0.5 Å^{-1}. In the present cases the theoretical scattering curves from *CRYSOL* agree well with the experimental scattering curves up to $q \simeq 0.7$ Å^{-1}. Then we used the Debye equation, equation (1), for calculating the scattering curves above $q \simeq 0.7$ Å^{-1}. In Fig. 3 the theoretical scattering curves obtained by using both the Debye equation and the *CRYSOL* program are shown to fit to the full experimental scattering curves in Fig. 1. In the case of myoglobin and lysozyme, we can recognize relatively good agreements between the theoretical and experimental scattering curves over the whole q range in spite of the presence of the minor disagreements in the high-angle scattering region ($q > 1.5$ Å^{-1}). As already mentioned (Pickover & Engelman, 1982), the minor disagreements at $q > 1$ Å^{-1} can probably be attributed to a certain multiplicity of protein structures in solutions. Namely, intramolecular motions or local displacements of a protein entrapped in a periodic crystal lattice potential would be rather restricted in a defined space. Thus, such thermal local motions of a protein in a solution would be much enhanced compared with those in a crystalline state, which would contribute more evidently to the smearing of the experimental scattering curves at $q > 1$ Å^{-1}. Another probable reason could be the

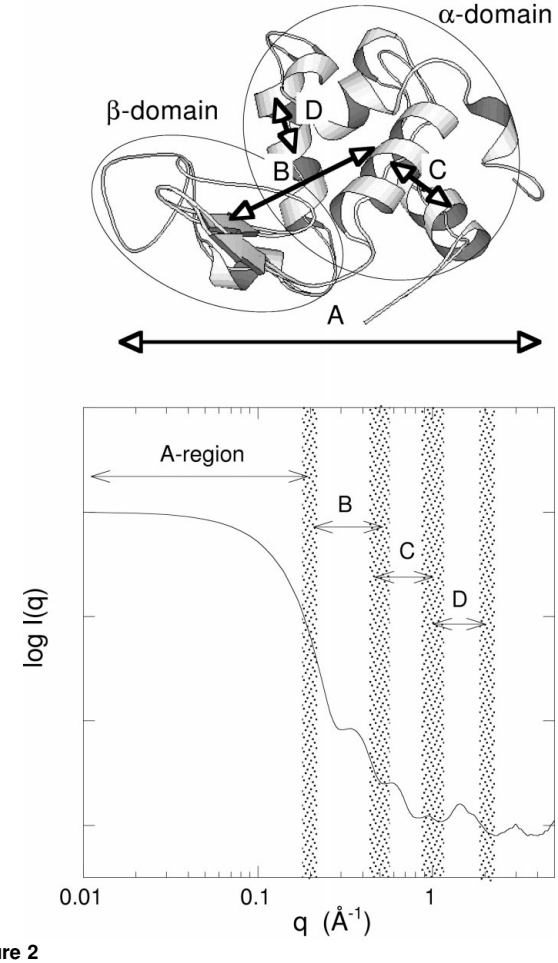

Figure 2
The correspondence of the theoretical solution scattering curve of hen egg-white lysozyme to typical intramolecular distances in its three-dimensional structure is shown schematically. The theoretical scattering curve was obtained by the Debye equation. The three-dimensional picture is presented using the *MolScript* program.

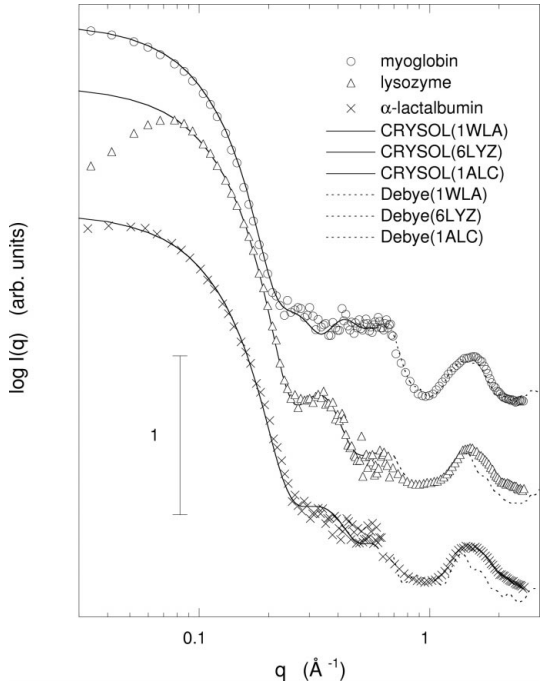

Figure 3
Comparison of experimental and theoretical scattering curves over the full q range measured. The experimental scattering curves in Fig. 1 were fitted by the theoretical scattering curves using both the Debye equation and the *CRYSOL* program.

atomic coordinates used for the calculation. Most of the PDB files lack the atomic coordinates of H atoms and hydrated water molecules surrounding protein surfaces. A recent cryogenic X-ray crystal structure analysis (Nakasako, 1999) shows clearly that a protein surface is covered by hydrated water molecules which form aggregates partly brought about by large-scale networks of hydrogen bonds. Such large-scale networks of hydrated water molecules localized at a protein surface would change not only the radius of gyration of the protein but would also modulate the scattering curve above $q \simeq 1$ Å^{-1}.

In the case of α-lactalbumin, the deviation between the theoretical and experimental scattering curves in the high-angle scattering region ($q > 1.5$ Å^{-1}) is very evident. α-Lactalbumin takes a tertiary structure similar to that of lysozyme; however, the unfolding-and-folding transitions of those proteins are known to be quite different. Namely, α-lactalbumin has been studied well to take a transition intermediate, a so-called molten globule state (Peng *et al.*, 1995; Radford & Dobson, 1995; Kuwajima, 1996). In addition, the thermal transition processes of those proteins observed by the solution X-ray scattering also significantly differ (Hirai *et al.*, 1998). Such differences between those proteins in the unfolding-and-folding process and the thermal stability would be reflected in the high-angle scattering curves ($q > 1.5$ Å^{-1}) as the evident deviation between the theoretical (at crystal state) and experimental (at solution state) scattering curves.

4. Conclusion

Combined with our previous reports, we can conclude as follows. The wide-angle solution scattering data obtained from the proteins in the q range 0.2–2.5 Å^{-1} with high statistics by using a high-intensity X-ray source are expected to provide further information on intramolecular structures such as intramolecular domain structures and their correlations, and secondary structures and their periodicities and packings. As mentioned in the *Introduction*, there are only a few experimental results treating wide-angle solution scattering data of proteins up to $\sim$1 Å^{-1} (Pickover & Engelman, 1982; Stuhrmann, 1973). Those wide-angle solution scattering data were used to develop modelling calculation methods of solution scattering curves of biological macromolecules, especially of proteins (Fedorov *et al.*, 1974, 1976; Fedorov & Alexander, 1978; Müller, 1983). In spite of the early studies of solution scattering curve calculations of proteins over wide-angle scattering regions, unfortunately many efforts on theoretical scattering curve calculation have been focused on small-angle scattering data analyses to determine overall solute particle shapes (Chacón *et al.*, 1998). This would be simply due to the lack of high-angle experimental scattering data with high statistics.

The present scattering data at high q seem to be smeared and lack some fine structure obtained by the theoretical calculation using the Debye equation. As described above, this may be attributed to hydrated water on the protein surface, intramolecular local motion, the difference between the physicochemical solvent conditions of solution and crystal, or temperature, and so on. Further experiments on various proteins under various conditions would be necessary to perform detailed analyses of high-angle scattering profiles from proteins in solutions, and experimental scattering profiles can be compared with theoretical ones from molecular dynamics simulations. Although the present report treats only a few proteins under limited conditions, we think that the synchrotron radiation wide-angle X-ray solution scattering method would still be a powerful and promising tool for studying structural hierarchy and the function of various proteins, especially a mechanism of folding-and-unfolding of proteins.

This work was performed under the approval of the JASRI Program Advisory Committee (Proposal No. 1999B0229-NL and 2001-A0001-NL).

References

Acharya, K. R., Stuart, D. I., Walker, N. P., Lewis, M. & Phillips, D. C. (1989). *J. Mol. Biol.* **208**, 99–127.
Arai, S. & Hirai, M. (1999). *Biophys. J.* **76**, 2192–2197.
Bosio, L., Teixeria, J. & Bellissent-Funel, M.-C. (1989). *Phys. Rev. A*, **39**, 6612–6613.
Chacón, P., Morán, F., Díaz, J. F., Pantos, E. & Andreu, J. M. (1998). *Biophys. J.* **74**, 2760–2775.
Debye, P. (1915). *Ann. Phys.* **46**, 809–823.
Diamond, R. (1974). *J. Mol. Biol.* **82**, 371–391.
Fedorov, B. A. & Alexander, A. I. (1978). *J. Appl. Cryst.* **11**, 473–477.
Fedorov, B. A., Kröber, R., Damashun, G. & Ruckpaul, K. (1976). *FEBS Lett.* **65**, 92–95.
Fedorov, B. A., Ptitsyn, O. B. & Voronin, L. A. (1974). *J. Appl. Cryst.* **7**, 181–186.
Feigin, L. A. & Svergun, D. I. (1987). *Structure Analysis by Small Angle X-ray and Neutron Scattering*, edited by G. W. Taylor, pp. 59–182. New York: Plenum.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
Hirai, M., Arai, S. & Iwase, H. (1999). *J. Phys. Chem. B*, **103**, 549–556.
Hirai, M., Arai, S. & Iwase, H. (2000). *Thermochim. Acta*, **344**, 95–102.
Hirai, M., Arai, S., Iwase, H. & Takizawa, T. (1998). *J. Phys. Chem. B*, **102**, 1308–1313.
Hirai, M., Arai, S., Takizawa, T., Yabuki, S. & Sano, Y. (1997). *Phys. Rev. B*, **55**, 3490–3496.
Hirai, M., Hirai, R. K., Hirai, T. & Ueki, T. (1993). *Eur. J. Biochem.* **215**, 55–61.
Hirai, M., Iwase, H. & Hayakawa, T. (1999). *J. Phys. Chem. B*, **103**, 10136–10142.
Hirai, M., Takizawa, T., Yabuki, S. & Hayashi, K. (1996). *J. Chem. Soc. Faraday Trans.* **92**, 4533–4540.
Kuwajima, K. (1996). *FASEB J.* **10**, 102–109.
Maurus, R., Overall, C. M., Bogumil, R., Luo, Y., Mauk, A. G., Smith, M. & Brayer, G. D. (1997). *Biochim. Biophys. Acta*, **1341**, 1–13.
Miura, K., Kawamoto, M., Inoue, K., Yamamoto, M., Kumasaka, T., Sugiura, M., Yamano, A. & Moriyama, H. (2000). SPring-8 User Experiment Report, Vol. 4, p. 168. SPring-8, Hyogo, Japan.
Müller, J. J. (1983). *J. Appl. Cryst.* **16**, 74–82.
Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia, C. (1995). *J. Mol. Biol.* **247**, 536–540.
Nakasako, M. (1999). *J. Mol. Biol.* **289**, 547–564.
Peng, Z., Wu, P. L. & Kim, P. S. (1995). *Biochemistry*, **34**, 3248–3252.
Perutz, M. F., Fermi, G., Poyart, C., Pagnier, J. & Kister, J. (1993). *J. Mol. Biol.* **233**, 536–545.
Pickover, C. A. & Engelman, D. M. (1982). *Biopolymers*, **21**, 817–831.
Pilz, I. (1982). *Small Angle X-ray Scattering*, edited by O. Glatter & O. Kratky, pp. 239–293. London: Academic.
Radford, S. & Dobson, C. (1995). *Philos. Trans. R. Soc. B*, **348**, 17–25.
Stuhrmann, H. B. & Miller, A. (1978). *J. Appl. Cryst.* **11**, 325–345.
Svergun, D. I., Barberato, C. & Koch, M. H. J. (1995). *J. Appl. Cryst.* **28**, 768–773.
Svergun, D. I., Richard, S., Koch, M. H. J., Sayers, Z., Kuprin, S. & Zaccai, G. (1998). *Proc. Natl. Acad. Sci. USA*, **95**, 2267–2272.
Stuhrmann, H. B. (1973). *J. Mol. Biol.* **77**, 363–369.
Trewhella, J., Gallagher, S. C., Krueger, J. K. & Zhao, J. (1998). *Sci. Progr.* **81**, 101–122.
Tsukada, H. & Blow, D. M. (1985). *J. Mol. Biol.* **184**, 703–711.
Wlodawer, A., Svensson, L. A., Sjolin, L. & Gilliland, G. L. (1988). *Biochemistry*, **27**, 2705–2717.

 Received 30 March 2001 • Accepted 12 April 2002