ELSEVIER

Biophysical Chemistry 95 (2002) 235–252

Biophysical Chemistry

www.elsevier.com/locate/bpc

Contribution of solvent water to the solution X-ray scattering profile of proteins

Yasutaka Seki[a], Tadashi Tomizawa[a], Nikolay N. Khechinashvili[b], Kunitsugu Soda[a,*]

[a]*Department of Bioengineering, Nagaoka University of Technology, Kamitomioka-cho, Nagaoka, Niigata 940-2188, Japan*
[b]*Institute of Cell Biophysics, Russian Academy of Sciences, 142292 Pushchino, Moscow Region, Russia*

Received 8 May 2001; accepted 1 November 2001

Abstract

A theoretical framework is presented to analyze how solvent water contributes to the X-ray scattering profile of protein solution. Molecular dynamics simulations were carried out on pure water and an aqueous solution of myoglobin to determine the spatial distribution of water molecules in each of them. Their solution X-ray scattering (SXS) profiles were numerically evaluated with obtained atomic-coordinate data. It is shown that two kinds of contributions from solvent water must be considered to predict the SXS profile of a solution accurately. One is the excluded solvent scattering originating in exclusion of water molecules from the space occupied by solutes. The other is the hydration effect resulting from formation of a specific distribution of water around solutes. Explicit consideration of only two molecular layers of water is practically enough to incorporate the hydration effect. Care should be given to using an approximation in which an averaged electron density distribution is assumed for the structure factor because it may predict profiles considerably deviating from the correct profile at large *K*.

Keywords: Solution X-ray scattering; Scattering profile; Partial molecular scattering; Hydration structure; Hydration effect; Myoglobin

1. Introduction

A natural protein forms a unique native structure under the physiological condition to carry out its specific biological function. It is beyond discussion that 3D atomic-coordinate data of a protein are essential for considering the molecular mechanism of its function and structural stabilization. It is because the assumption holds at least approximately for most water-soluble proteins that the crystal structure is basically maintained in aqueous solution. However, the thermodynamic stability of their native structure is as marginal as approximately 10 kcal/mol [1]. As a result of it, most of the proteins may undergo a structural transition into some non-native state due to a small change in solution conditions. Hence, to study the design principle of protein structure, it is essential to get information on the native structure under the physiological condition, and the non native structure

*Corresponding author. Fax: +81-258-47-9424.
E-mail address: soda@vos.nagaokaut.ac.jp (K. Soda).

0301-4622/02/$ - see front matter
PII: S0301-4622(01)00260-5

under non-physiological conditions [2–16], as well as the crystal structure.

Many experimental methods have been used to investigate the solution structure of proteins. Spectroscopic probes as circular dichroism, fluorescence and nuclear magnetic resonance yield information mainly on the local structure of proteins. Complementary to them, solution X-ray scattering (SXS) along with solution neutron scattering yields information on the molecular size and shape of solute molecules [17,18]. (Though SXS has conventionally been called small angle X-ray scattering, it will be called here as above to contrast with X-ray crystallography.) This characteristic of the SXS method has successfully been employed for the structural analysis of not only small proteins but also many biological structures such as ribosome and viruses. With the recent prevalence of the strong X-ray source of synchrotron radiation, the SXS method is expected to play an important role for studying the solution structure of proteins more precisely.

Dependence on scattering-angle of the intensity of X-ray scattered from a medium has been called the scattering profile. Experimentally, the contribution of a solute molecule to the scattering profile of a solution is determined from the difference between the scattering profiles of solution and solvent divided by solute concentration and extrapolated to infinite dilution. This increment of the scattering profile of a solution due to introducing a solute molecule into solution will be hereafter called the partial molecular scattering (PMS) profile after the terminology of the partial molecular quantities. X-Ray scattering from a solution arises from the spatial inhomogeneity of electron density in it. In an aqueous protein solution, not only the atoms in protein molecules but also water molecules surrounding them contribute to the inhomogeneity. Various atomic groups on the surface of a protein make specific interactions with nearby water molecules [19,20]. They have a spatial distribution different from that of bulk water, which is generally called the hydration structure [21,22]. This implies that hydration water can also contribute to the PMS profile. Hence, it will be possible to extract information on the hydration structure as well as the 3D structure of proteins in solution from analysis of PMS profiles.

For reproducing experimental profiles, several methods have been reported to calculate the PMS profile of proteins with the contribution of solvent water included in various ways. The cube method [23,24], the improved cube method [25,26] and the surface integration method [27] were devised to incorporate the contrast effect of solvent. A method for rapidly calculating the PMS profile from atomic-coordinate data on solute and solvent was described by Lattman [28]. Svergun et al. [29] developed a computer program, CRYSOL, for evaluating the SXS profile of biomacromlecules approximately taking into account the hydration contribution with a simple hydration shell model. Experimental observations of protein hydration by SXS have also been reported [30,31]. In spite of these works, it has not yet been fully examined, theoretically as well as experimentally, how the hydration structure of a protein is reflected on its PMS profile.

The purpose of this study is to clarify the contribution of solvent water to the PMS profile of a solute protein by reproducing the actual experimental process as closely as possible in the computer. To solve the problem, a theoretical framework is first presented for decomposing an SXS profile into six different contributions. Molecular dynamics (MD) simulations were made on pure water and an aqueous solution of myoglobin containing a protein molecule and water. Average contributions of the respective scattering components to the PMS profile were evaluated numerically with applying the presented formulas to atomic-coordinate data. It is confirmed from SXS measurements of myogobin that the calculated profile coincides well with the experimental one. Based on the results, it is discussed how solvent water contributes to the PMS profile of protein.

2. Method of analysis

2.1. Theoretical basis

Let us consider a dilute aqueous solution where a protein molecule is dissolved in a large amount of water. If the protein concentration is sufficiently

low, the average electron density of the solution can be assumed equal to that of pure water denoted as $\bar{\rho}$. Denoting the electron density at position r in a scattering medium, solution or solvent, as $\rho(r)$, we have the structure factor $F(K)$ of the medium for scattering vector K as follows:

$$F(K) = \int_V \rho(r)\exp(iK\cdot r)\mathrm{d}r = \int_V \left(\rho(r)-\bar{\rho}\right)\exp(iK\cdot r)\mathrm{d}r + (2\pi)^3\bar{\rho}\delta(K) \quad (1)$$

Let us introduce the spatial fluctuation of electron density $\delta\rho(r)$ by

$$\delta\rho(r) \equiv \rho(r) - \bar{\rho} \quad (2)$$

By the definition of $\bar{\rho}$, the average of $\delta\rho(r)$ must be zero. The structure factor $F(K)$ of the medium is expressed as

$$F(K) = \int_V \delta\rho(r)\exp(iK\cdot r)\mathrm{d}r, \quad (K \neq 0) \quad (3)$$

We will divide the total volume of the solution (u) and pure solvent (v) into the following three regions, respectively (Fig. 1):

$$\mathrm{u} = \mathrm{p}\cup\mathrm{h}\cup\mathrm{b}, \quad \mathrm{v} = \bar{\mathrm{p}}\cup\bar{\mathrm{h}}\cup\bar{\mathrm{b}} \quad (4)$$

where 'p', 'h' and 'b' stand for the 'protein', the 'hydration' water and the 'bulk' water regions, respectively. The 'h' region will be taken large enough for the statistical properties of water in the 'b' region to be assumed practically the same as those of bulk water. Similarly, we will define the regions $\bar{\mathrm{p}}$, $\bar{\mathrm{h}}$ and $\bar{\mathrm{b}}$ in pure solvent that have the same shape as p, h and b, respectively. Division of pure solvent into the three regions is made only conceptually and there is no difference among the properties of water in these regions. The fluctuations of electron density, $\delta\rho_{\mathrm{u}}(r)$ and $\delta\rho_{\mathrm{v}}(r)$, in solution and solvent are expressed as

$$\delta\rho_{\mathrm{u}}(r) = \delta\rho_{\mathrm{p}}(r) + \delta\rho_{\mathrm{h}}(r) + \delta\rho_{\mathrm{b}}(r) \quad (5)$$

$$\delta\rho_{\mathrm{v}}(r) = \delta\rho_{\bar{\mathrm{p}}}(r) + \delta\rho_{\bar{\mathrm{h}}}(r) + \delta\rho_{\bar{\mathrm{b}}}(r) \quad (6)$$

where $\delta\rho_\lambda(r)$ ($\lambda = \mathrm{p}, \mathrm{h}, \mathrm{b}, \bar{\mathrm{p}}, \bar{\mathrm{h}}, \bar{\mathrm{b}}$) can have a non-zero value only when the position r is in the region λ. Corresponding to this decomposition, the structure factors of solution and solvent, $F_{\mathrm{u}}(K)$ and $F_{\mathrm{v}}(K)$, consist of the three contributions, respectively, described as

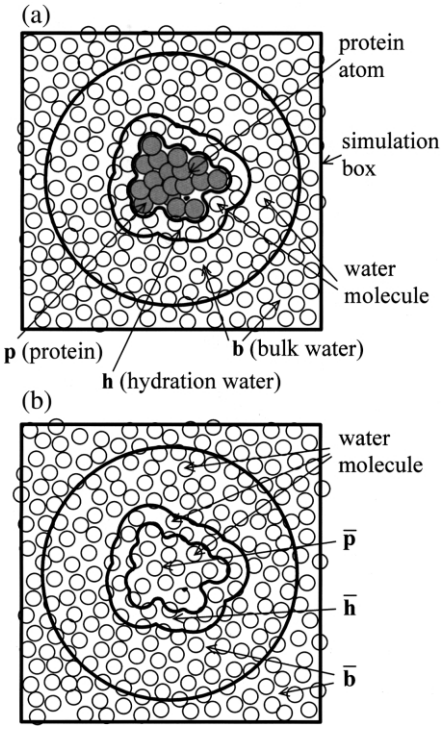

Fig. 1. Illustration of dividing (a) the solution volume into the three regions of 'p' (protein), 'h' (hydration water) and 'b' (bulk water), and (b) the solvent volume into the corresponding regions of '$\bar{\mathrm{p}}$', '$\bar{\mathrm{h}}$' and '$\bar{\mathrm{b}}$'. The circle and the square represent the scattering sphere and the simulation box, respectively.

$$F_{\mathrm{u}}(K) = F_{\mathrm{p}}(K) + F_{\mathrm{h}}(K) + F_{\mathrm{b}}(K) \quad (7)$$

$$F_{\mathrm{v}}(K) = F_{\bar{\mathrm{p}}}(K) + F_{\bar{\mathrm{h}}}(K) + F_{\bar{\mathrm{b}}}(K) \quad (8)$$

$$F_\lambda(K) = \int_V \delta\rho_\lambda(r)\exp(iK\cdot r)\mathrm{d}r, \quad \left(\lambda = \mathrm{p}, \mathrm{h}, \mathrm{b}, \bar{\mathrm{p}}, \bar{\mathrm{h}}, \bar{\mathrm{b}}\right) \quad (9)$$

Denoting the atomic structure factor of the ith atom at position $r_{\lambda,i}$ in the region λ as $f_{\lambda,i}(K)$ and making use of Eq. (2) and Eq. (3), we have the structure factor of the region λ as

$$F_\lambda(K)=\sum_i f_{\lambda,i}(K)\cdot\exp(iK\cdot r_{\lambda,i}) - \bar{\rho}\int_{V\lambda}\exp(iK\cdot r)\mathrm{d}r \quad (10)$$

The values of $f_{\lambda,i}(K)$'s are obtained from literature [32]. The volume integral in the second term of the right hand side of Eq. (10) represents the contrast effect between solute and solvent. It can be readily evaluated with a surface integral over the interface between the region λ and the contiguous region using Gauss's theorem [27]. The solution scattering profile $I_u(K)$ and the solvent scattering profile $I_v(K)$ are given by

$$I_u(K)=|F_u(K)|^2=\sum_{\lambda,\mu=p,h,b}F_\lambda(K)\cdot F_\mu^*(K), \quad (11)$$

$$I_v(K)=|F_v(K)|^2=\sum_{\lambda,\mu=\bar{p},\bar{h},\bar{b}}F_\lambda(K)\cdot F_\mu^*(K), \quad (12)$$

respectively. Let us define functions $I_{\lambda\lambda}(K)$ and $I_{\lambda\mu}(K)$ by the following expressions as

$$I_{\lambda\lambda}(K)=|F_\lambda(K)|^2=F_\lambda(K)\cdot F_\lambda^*(K), \quad (13)$$

$$I_{\lambda\mu}(K)=F_\lambda(K)\cdot F_\mu^*(K)+F_\lambda^*(K)\cdot F_\mu(K) =I_{\mu\lambda}(K),\quad(\lambda\neq\mu). \quad (14)$$

The functions $I_{\lambda\lambda}(K)$ and $I_{\lambda\mu}(K)$ represent contributions from the interference effect of X-ray scattered from the atoms within the region λ, and that from the atoms in two regions λ and μ, respectively. Thus, each of $I_u(K)$ and $I_v(K)$ is decomposed into six scattering components as

$$I_u=I_{pp}+I_{hh}+I_{bb}+I_{ph}+I_{pb}+I_{hb}, \quad (15)$$

$$I_v=I_{\bar{p}\bar{p}}+I_{\bar{h}\bar{h}}+I_{\bar{b}\bar{b}}+I_{\bar{p}\bar{h}}+I_{\bar{p}\bar{b}}+I_{\bar{h}\bar{b}}, \quad (16)$$

where the variable K is abbreviated in both equations. In our case, the partial molecular scattering (PMS) profile, $I_{pms}(K)$, is simply given by

$$I_{pms}(K)=I_u(K)-I_v(K) =\Delta I_{pp}+(\Delta I_{ph}+\Delta I_{hh}) +(\Delta I_{pb}+\Delta I_{hb}+\Delta I_{bb}) \quad (17)$$

The function $\Delta I_{\xi\eta}(K)$ defined by the following equation:

$$\Delta I_{\xi\eta}(K)\equiv I_{\xi\eta}(K)-I_{\bar{\xi}\bar{\eta}}(K),$$
$$(\xi,\ \eta=p,\ h,\ b\quad \bar{\xi},\ \bar{\eta}=\bar{p},\ \bar{h},\ \bar{b}), \quad (18)$$

represents the difference between the interference effects of X-ray scattered from the regions ξ and η in solution and those from the corresponding regions $\bar{\xi}$ and $\bar{\eta}$ in pure solvent.

As seen from Eq. (17), the PMS profile is decomposed into six contributions coming from the different combinations of two regions taken out of the three different ones. Grouping these contributions into three as shown in Eq. (17), we will briefly describe their characteristics:

(a) ΔI_{pp}

This is the major part of $I_{pms}(K)$ and is the difference between the intensity of X-ray scattered from the protein atoms in the 'p' region, I_{pp}, and that from the water molecules in the corresponding '$\bar{p}$' region, $I_{\bar{p}\bar{p}}$. The latter will be called the excluded solvent scattering as it corresponds to the scattering contribution from solvent molecules in solution excluded with introduction of a solute molecule. This contribution will vanish under the approximation that solvent is a continuous medium with uniform electron density. In fact, it has been neglected in many of the conventional analyses of SXS profiles. When the molecular nature of solvent is considered explicitly, however, it should have a non-zero contribution. Evaluation of its magnitude described in Section 3.2 will show that it can never be neglected and is more significant at larger K.

(b) $\Delta I_{ph}+\Delta I_{hh}$

These are contributions containing the interference effect of the atoms in the regions 'p' and 'h', and that within the region 'h', respectively. Generally, the electron density of a protein molecule is higher than that of bulk water. Besides it, there are various atomic groups with different polarities on the surface of a protein. Water molecules interact specifically with protein surface groups and have in general a spatial distribution different from that of bulk water. It will be naturally expected that both ΔI_{ph} and ΔI_{hh} have a significantly non-zero value. Hence, these contributions will reflect the net hydration effect of the solute protein.

(c) $\Delta I_{pb}+\Delta I_{hb}+\Delta I_{bb}$

It has been suggested from many works that the spatial correlation length of bulk water is shorter than 1 nm. As the regions 'p' and 'b' are separated by the hydration region 'h', the interference effect between the regions '$\bar{p}$' and '$\bar{b}$' will be small or $I_{\bar{p}\,\bar{b}} \approx 0$. Therefore, ΔI_{pb} will be contributed mainly from I_{pb}. As the magnitude of ΔI_{hb} depends on how much the spatial correlation of hydration water and bulk water differs from that within bulk water, it will reflect the structural difference between water in the two regions. According to similar consideration, the difference in statistical properties of water in the regions 'b' and '$\bar{b}$' will be small by the definition of 'b'. Therefore, we can safely assume that $\Delta I_{bb} \approx 0$ and its contribution to $I_{pms}(K)$ is negligible. It is readily expected that the more dilute solution has the larger relative magnitude of I_{bb} in I_u. The purpose of subtracting the solvent scattering $I_v(K)$ from the solution scattering $I_u(K)$ in the analysis of experimental SXS data is to cancel I_{bb} with $I_{\bar{b}\,\bar{b}}$ and realize the situation of $\Delta I_{bb} \approx 0$.

2.2. Molecular dynamics (MD) simulation

To estimate solvent contribution to the SXS profile of an aqueous protein solution, it is necessary to evaluate magnitudes of the respective terms in Eq. (17). For this purpose, molecular dynamics (MD) calculations have been made on pure water and an aqueous solution of myoglobin. The ROAR ver. 2.0 of AMBER ver. 6.0 and AMBER96 were used for the MD simulation program and the force field, respectively [33,34]. Non-bonded interactions are evaluated with a cut-off length of 0.8 nm, where all the atoms in water molecules and amino acid residues are taken into account when any of the atoms in them is located within the cut-off distance. To avoid a possible increase in the amount of numerical calculation due to incorporating long range Coulomb interactions and the effect of large errors due to inadequate approximations, all ionic groups of myoglobin are neutralized to have no net charge. The internal water molecules found in the crystal structure are incorporated also in the solution structure taken for MD simulation.

The MD simulation of pure water was made as follows: An isothermal–isobaric MD calculation was done for 16,100 TIP3P-model water molecules in a cube under the periodic boundary condition. The temperature and the pressure were set to 298.15 K (25 °C) and 1.01325×10^5 Pa (1 atm), respectively. The time step Δt was taken to be 1 fs. After an equilibrating simulation of 20 ps, a regular MD calculation was made for 80 ps, from which result the average volume of the simulation box was determined. The edge length of the cube having a volume equal to the average volume was found to be 7.867 nm. With the edge length of the cube holding to this value, an isothermal–isochoric MD calculation was performed for 500 ps. The above cube size corresponds to the water density of 0.98916 g/ml, which is 0.79% lower than the value of real water at 25 °C, 0.99704 g/ml. The average pressure of water was found 2.888 MPa (28.50 atm). The difference between this value and the expected value of 1 atm will be caused by a small difference between the fixed volume and the true average volume at 1 atm. It is well known that accurate determination of the average pressure is very difficult for liquid simulation, and deviation of this magnitude of the average pressure will not affect significantly the statistical properties of water molecules needed for the calculation of its SXS profile. Atomic coordinates of the 16,100 water molecules were stored every 1 ps and 500 sets of coordinate data were obtained.

Next, an aqueous solution of myoglobin was 'prepared' in computer. It consists of one protein molecule and 15,629 TIP3P-model water molecules. All atoms in myoglobin are fixed to their position in crystal structure provided by PDB (Protein Data Bank). A myoglobin molecule is placed so that its center of gravity coincides with the center of the simulation box. In a similar manner to the case of pure water, the edge length of the box was determined as 7.900 nm. With the fixed box size, an isothermal–isochoric MD simulation was done for 500 ps. The average pressure of the solution was found to be −5.833 MPa (−57.57 atm). The partial specific volume of myoglobin determined using the density of bulk water estimated from the MD simulation of pure water

Table 1
Atomic radii taken for the calculation of SXS profiles [37]

Atom type	Radius/nm
sp^3 Carbon	0.187
sp^2 Carbon	0.176
Sulfur	0.185
sp^2 Nitrogen	0.165
Oxygen	0.140
sp^3 Charged nitrogen	0.150
sp^2 Charged nitrogen	0.165
Charged oxygen	0.140

is 0.7474 ml/g, which agrees well with its experimental value.

2.3. Numerical calculation of SXS profiles

Calculation of SXS profiles using the result of MD simulations was performed based on the formulas described in Section 2.1. Details of the procedure will be reported elsewhere. Here, only the method of dividing a protein solution into the three regions, 'p', 'h' and 'b', will be described. If the regions 'p', 'h' and 'b' are determined, then dividing pure water solvent into the three regions '$\bar{p}$' '$\bar{h}$' and '$\bar{b}$' can be made straightforward. We will call the boundary between the regions 'p' and 'h' in solution 'the ph boundary' and the boundary between the regions '$\bar{h}$' and '$\bar{b}$' in solvent 'the $\bar{hb}$ boundary', and so on.

Here, each atom of a myoglobin molecule is assumed to be modeled by a hard sphere with an intrinsic radius and a water molecule by a sphere with radius $r_w = 0.14$ nm. With this assumption, the molecular volume of a protein is definitely determined as the volume of the region inaccessible to any part of the probe water sphere [35]. The molecular surface is defined as the boundary between the accessible and inaccessible regions. Similarly, the excluded volume and the accessible surface can also be defined if the above 'any part' is replaced with 'the center' [36]. The values of atomic radii are taken from Chothia et al. [37] (Table 1).

In the calculation of SXS profiles, only X-rays scattered from the molecules located inside the sphere of radius $R_s = 3.80$ nm were explicitly taken into account, where the center of the sphere coincides with the center of gravity of the protein. This sphere will be, hereafter, called the scattering sphere. The region outside the scattering sphere is approximately replaced with a continuous medium having the same uniform electron density as that of bulk water. This approximation is made considering that behavior of the water molecules near the corner of the simulation box tend to be affected by water molecules in the neighboring boxes and their statistical properties might slightly deviate from those of bulk water.

First, we will consider the hb boundary. The hydration region 'h' needs to be defined as the region whose contribution to the SXS profile differs significantly from that of bulk water. It is expected to be a narrow region near the molecular surface. For candidates of the hb boundary, we have generated many surfaces each of which is equally t distant from the molecular surface of myoglobin (Fig. 2). The position of the true hb boundary at $t = t_b$ was estimated from the dependence on t of the contribution from the assumed 'h' region to $I_{pms}(K)$. Details for generating the equidistant surfaces are described in the Appendix.

Next, we will consider the ph boundary (Fig. 2). A simplest definition of the ph boundary may be the molecular surface. However, the molecular volume or the volume inside the molecular surface

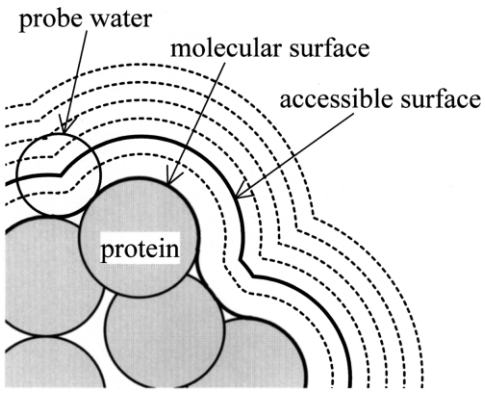

Fig. 2. Surfaces enclosing the solute protein molecule, each of which is generated equidistant from the molecular surface of the protein. Details of the method for generating the surfaces are described in the Appendix.

of a protein is in general smaller than its partial molecular volume V_p. As V_p of a protein is directly related to the forward scattering intensity, the ph boundary should be chosen so that the volume inside the boundary is equal to V_p. Therefore, similarly to the case of the pb boundary, we have taken the ph boundary so that it is separated $t=t_h$ from the molecular surface and it encloses the volume equal to V_p of myoglobin.

The value of V_p was determined from the result of MD simulation as follows: First, we evaluated the volume $\bar{V}_{b0}$ of the region with a distance between 0.8 and 1.2 nm from the accessible surface of protein and the average number $\bar{N}_{b0}$ of water molecules in the region. Water molecules in the region are expected to have a spatial distribution similar to that of bulk water. The partial molecular volume $\bar{v}_{b0}$ of bulk water is estimated as

$$\bar{v}_{b0}=\bar{V}_{b0}/\bar{N}_{b0} \quad (19)$$

The partial molecular volume V_p of a protein is defined by the increase in solution volume when a protein molecule is introduced into solution. The value of V_p is obtained from the volume of the scattering sphere, $V_s=(4\pi/3)R_s^3$, and the average number of water molecules, N_w, in the sphere as

$$V_p=V_s-N_w\bar{v}_{b0} \quad (20)$$

As seen from the definition of the ph boundary, the region inside the boundary can be regarded as the region that the protein molecule occupies virtually or the 'p' region. The ph boundary surface may, therefore, be called the effective molecular surface. The value of t_h for myoglobin was obtained as 0.0325 nm. It was confirmed that the partial molecular volume thus determined agrees nearly with the experimental value.

With the above definition of two boundaries, the 'h' region is a shell-like region bound inward by the ph boundary or the effective molecular surface, and bound outward by the pb boundary. Any point inside the 'h' region is at a distance between t_h and t_b from the molecular surface. Thus, determination of the real hydration region is reduced to the problem of determining the distance t_b of the hb boundary surface.

As described above, in the calculation of SXS profiles, only the X-ray scattering from the solute protein and water molecules in the scattering sphere of radius R_s is taken into account explicitly. In other word, only the contrast effect is considered for the contribution from the region outside the sphere by assuming it as a continuous medium with the same electron density as bulk water. Most of the MD calculations have been made assuming as $R_s=3.80$ nm. With this size of the scattering sphere, even the protein atom most distant from the protein center is separated from the surface of the scattering sphere by a distance of three water layers. To examine effects of the finite size of the scattering sphere, calculations were made also for three values of $R_s=3.95$, 3.50 and 3.00 nm.

Scattering profiles of solvent water were calculated applying equations in Section 2.2 to 500 sets of coordinate data for the water molecules in the scattering sphere. Six scattering contributions from solvent were evaluated dividing formally the scattering sphere into the three regions of '$\bar{p}$', '$\bar{h}$' and '$\bar{b}$'.

2.4. Solution X-ray scattering experiment

To examine validity of our SXS calculations, SXS measurements were made on aqueous solutions of myoglobin. Freeze-dried samples of horse heart myoglobin for biochemical study were purchased from Nakarai Co. in Japan. Weighed protein was first dissolved in, and dialyzed against, Hepes buffer of pH 6.0. After concentration with ultra filtration, the protein solution was gel-filtrated with Sephadex G10 and diluted with Hepes buffer of pH 6.0 to serve for the SXS sample. SXS measurements were done using the beam line BL-10C of Photon Factory at Tsukuba in Japan. To see concentration dependence of scattering profiles, measurements were made for eight different concentrations: 2, 3, 5, 7.5, 10, 15, 20 and 40 mg/ml of myoglobin. A flow cell was used to make the actual irradiation time less than a second for minimizing the effect of X-ray radiation damage on protein molecules. Measurements on solution and solvent were made every 30 min in turn and X-ray scattering data were accumulated. Raw data were corrected for the effect of the change in time

of the incident X-ray intensity by monitoring it. The PMS profile of myoglobin was obtained from the difference between the accumulated SXS profiles of solution and solvent.

3. Results and discussion

3.1. Accuracy of the prediction for PMS profiles

Our simulation box has a micro volume that is only approximately 10 times the molecular volume of solute protein. It will be necessary to examine whether or not the calculated SXS profile for such a small system can correctly reproduce that for a real experimental system.

To investigate effects of the finite scattering volume in simulation, PMS profiles were calculated for three scattering spheres different in radius, $R_s = 3.95$, 3.50, 3.00 nm (Fig. 3). For brevity's sake, we will conveniently divide the whole K region into three as $K = 0 \sim 2$ nm^{-1}, $K = 2 \sim 6$ nm^{-1} and $K \geq 6$ nm^{-1}, and call the respective regions the small, medium and large K region. As shown in Fig. 3, differences among the PMS profiles for the three values of R_s are very small. This result shows that, if one of the three systems is a good approximation of the real solution, all the systems, including the smallest one of $R_s =$ 3.00 nm, serve for sufficiently good approximations. The scattering intensity of each solution is much higher than that of solvent in the small K region, but they are comparable to each other in the medium to large K regions. It means that, to analyze SXS profiles in the medium to large K region precisely, experimental data with high accuracy are needed on solvent as well as solution.

To illustrate that calculation reproduces the experiment fairly well, the calculated profile for the scattering sphere with $R_s = 3.80$ nm and the experimental profile are compared in Fig. 4 for myoglobin by Kratky plot. We can see that the calculated profile agrees well with the experimental one within the limited K range experimentally available to us. The mean square radius R_{sq} is estimated 1.720 nm from Guinier analysis of the calculated profile in the range of $0 \leq K^2 \leq 0.4$ nm^{-2}. The values of R_{sq} estimated for the three values of R_{sq} coincide with each other within

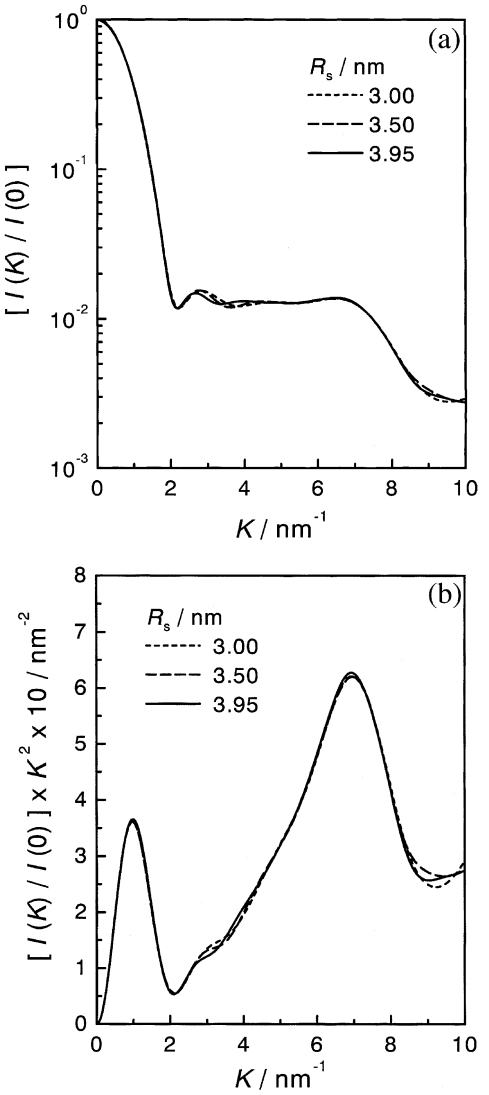

Fig. 3. Calculated PMS profiles of myoglobin for the three scattering spheres with different radii of $R_s = 3.00$, 3.50, 3.95 nm. (a) Semi-logarithmic plot and (b) Kratky plot.

$\pm 0.95\%$. Similar analysis of the experimental profile in the range of $0.13 \leq K^2 \leq 0.55$ nm^{-2} yields an estimate of $R_{sq} = 1.727$ nm. This result clearly shows that the experimental and calculated values of R_{sq} agree very well with each other. We conclude from those described above that, in spite of the difference in the systems dealt in calculation and experiment, calculation takes into account the

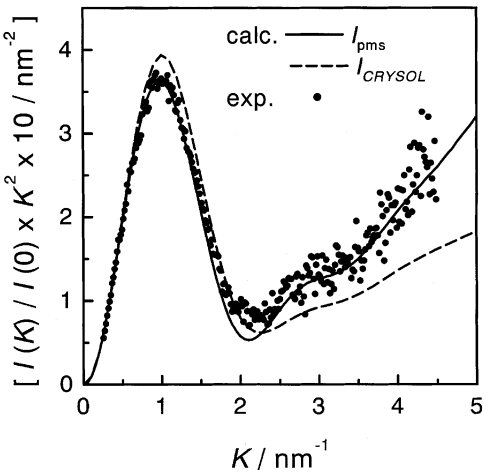

Fig. 4. Comparison of the experimental (•) and calculated (—) PMS profiles of myoglobin in aqueous solution at $T=298.15$ K. Calculation: $R_s=3.80$ nm; The number of TIP-3P water molecules are 15,629 and 16,100 for solution and solvent, respectively; The broken line (- - -) is an approximate profile calculated with the program CRYSOL developed by Svergun et al. [29]. Experiment: data in the small K region of $0\leq K\leq 1.5$ nm^{-1} were obtained from extrapolation of those for $C=2$, 3, 5, 7.5, 10, 20 and 40 mg/ml. Data at $K\geq 1.5$ nm^{-1} are those for $C=40$ mg/ml. In the larger K region, the experimental profile has no dependence on protein concentration detectable within experimental error.

essential aspects of the SXS experiment. The following calculations of the SXS profiles were made on the scattering sphere with radius $R_s=$ 3.80 nm, nearly the largest one among those accommodated in the simulation box.

3.2. Approximations and reproducibility of the PMS profile

Divided into terms as shown in Eq. (17), the PMS profile consists of six contributions, $\{\Delta I_{\xi\eta}\}$. Since each $\{\Delta I_{\xi\eta}\}$ is defined by $\Delta I_{\xi\eta}(K)\equiv I_{\xi\eta}(K)-I_{\bar{\xi}\bar{\eta}}(K)$ as in Eq. (18), the PMS profile $I_{pms}(K)$ can also be regarded composed of 12 contributions. If only a part of them are taken into account in the estimation of $I_{pms}(K)$, the resultant profile will be an approximation for $I_{pms}(K)$. The PMS profile evaluated with all the contributions included will be called the full PMS profile. Four approximations and their accuracy of prediction will be discussed here.

(1) Approximations of $I_{pms}\approx I_{pp}$ and $I_{pms}\approx \Delta I_{pp}=I_{pp}-I_{\bar{p}\bar{p}}$ (Fig. 5).In both approximations, the outside of the effective molecular surface is assumed a continuous medium with the same electron density as that of bulk water. Therefore, the contrast effect of solvent is incorporated but

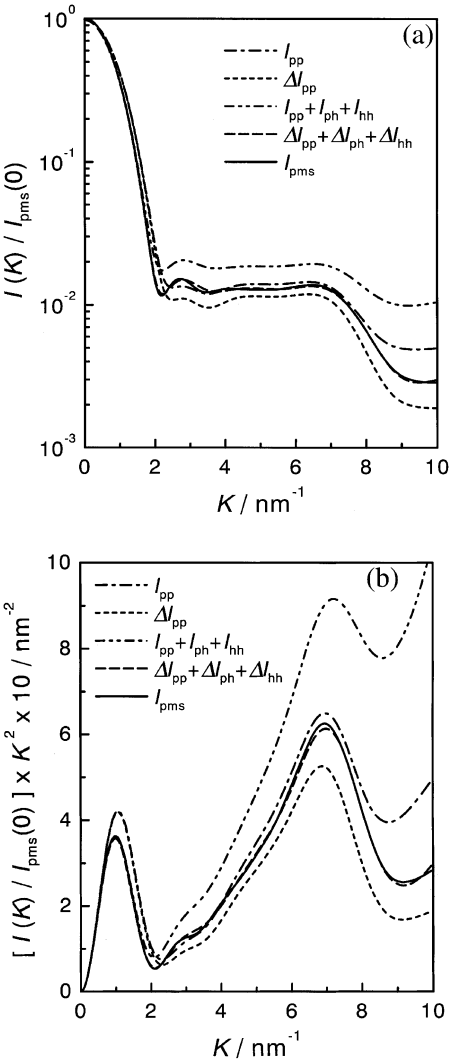

Fig. 5. Comparison of the full PMS profile and four approximate profiles of myoglobin. The pb boundary is assumed at distance $t_b=0.64$ nm from the molecular surface or at distance $d_b=0.50$ nm from the accessible surface. (a) Semi-logarithmic plot and (b) Kratky plot.

the hydration effect is not considered in both of them. The excluded solvent scattering is taken into account in the ΔI_{pp} approximation, while it is not considered in the I_{pp} approximation. Both approximations predict scattering intensity significantly higher than the full PMS profile in the small K region. It will be due to not taking into account the hydration effect in both of them. In the medium to large K region, the ΔI_{pp} approximation underestimates the scattering intensity, while the I_{pp} approximation predicts a profile near to or slightly higher than the full profile. This indicates that the effect of excluded solvent scattering becomes larger relatively at larger K. (2) Approximations of $I_{pms} \approx I_{pp} + I_{ph} + I_{hh}$ and $I_{pms} \approx \Delta I_{pp} + \Delta I_{ph} + \Delta I_{hh}$ (Fig. 5).

In these approximations, the hydration effect of the solute is considered almost correctly. The hb boundary is taken at distance $d_b = 0.50$ nm from the accessible surface or at distance $t_b = 0.64$ nm from the molecular surface. The outside of the boundary is assumed a medium with uniform electron density. As seen from the figure, both approximations reproduce very well the full PMS profile in the small K region. Comparing this result with those of the two approximations in (1), we can see that hydration water has a significant effect on the SXS profile in the small K region. In the larger K region, the $\Delta I_{pp} + \Delta I_{ph} + \Delta I_{hh}$ profile approximates the full profile quite well, while the $I_{pp} + I_{ph} + I_{hh}$ profile has large positive deviation from it. These results show that, to predict a PMS profile quantitatively over the whole K region: (a) it is necessary to take into account both scattering contributions from excluded solvent and hydration water; but (b) the hydration effect can be incorporated by considering explicitly only water molecules near the surface of protein. Later, we will discuss again quantitative analysis of the hydration effect.

Here, we will discuss the excluded volume scattering. As noted in Section 2.1, the principal purpose of using $I_{pms}(K) = I_u(K) - I_v(K)$ for the analysis of SXS profiles is to remove the large contribution of solvent molecules from the scattering profile of dilute solution. Excluded solvent scattering originates from the situation that whereas $I_u(K)$ has no solvent contribution from the 'p' region as solvent molecules are excluded from it, $I_v(K)$ includes solvent contribution from the '$\bar{p}$' region. Denoting the excluded solvent scattering in $I_v(K)$ as $I_{ev}(K)$, we can see from Eq. (16) that it is given by

$$I_{ev}(K) = I_{\bar{p}\,\bar{p}}(K) + I_{p\,\bar{h}}(K) + I_{p\,\bar{b}}(K) \tag{21}$$

To remove it from the original PMS profile, we need to suppose a hypothetical solvent v^* whose '$\bar{p}$' region is filled with not water but a uniform medium having the same electron density as bulk water or $\delta\rho_{\bar{p}}(r) = 0$. Then, the scattering profile, $I_{v*}(K)$, and the PMS profile, $I_{pms,\,1}(K)$, for this solvent are given by

$$\begin{aligned} I_{v*}(K) &= I_v(K) - I_{ev}(K) \\ &= I_{\bar{h}\,\bar{h}}(K) + I_{h\,\bar{b}}(K) + I_{b\,\bar{b}}(K) \end{aligned} \tag{22}$$

$$\begin{aligned} I_{pms,\,1}(K) &\equiv I_u(K) - I_{v*}(K) \\ &= I_{pms}(K) + I_{ev}(K) \\ &= I_{pp} + (I_{ph} + \Delta I_{hh}) + (I_{pb} + \Delta I_{hb}), \end{aligned} \tag{23}$$

respectively, where $\Delta I_{bb} = 0$ is assumed in Eq. (23). The PMS profile $I_{pms,1}(K)$ may be called the 'hydrated-molecule scattering' because it stands for the scattering from a solute molecule with its hydration effect included. With the volume fraction of the solute, ϕ, in solution known in advance, the quantity $I_{pms,\,2}(K)$ defined as

$$\begin{aligned} I_{pms,\,2}(K) &= I_u(K) - (1-\phi)\cdot I_v(K) \\ &= I_{pms}(K) + \phi I_v(K) \end{aligned} \tag{24}$$

has conventionally been used for analysis of experimental SXS profiles. We can see from comparison of Eq. (23) and Eq. (24) that the use of Eq. (24) assumes the following relation:

$$I_{ev}(K) = \phi \cdot I_v(K) \tag{25}$$

Equation 25 means that, in this analysis, the excluded solvent scattering is assumed to have the same profile as bulk water. It is apparent in this respect that the above relation never holds for any solute molecule strictly. The condition will approximately be satisfied by a molecule with globular shape so large that the surface effect in excluded solvent scattering can be neglected. Conversely, it will be a poor approximation for the molecules with linear conformation as unfolded proteins.

3.3. Decomposition of the PMS profile into six contributions

As shown in Eq. (17), the PMS profile $I_{pms}(K)$ consists of the six contributions, ΔI_{pp}, ΔI_{ph}, ΔI_{hh}, ΔI_{pb}, ΔI_{hb} and ΔI_{bb}. Relative magnitudes of these contributions will be evaluated for the case where the hb boundary is taken at distance $d_b = 0.50$ nm from the accessible surface as done in Section 3.2. Here, the X-ray scattering from water in the 'b' region inside the scattering sphere will also be considered explicitly. Decomposition of $I_{pms}(K)$ for myoglobin into the six components is shown in Fig. 6a,b.

The linear plot of Fig. 6a is normalized by the forward scattering intensity [18]:

$$I(0) = (N_{e,p} - \bar{\rho} V_p)^2 \quad (26)$$

where $\bar{\rho}$ is determined as the average electron density over the region with a distance between 0.8 and 1.2 nm from the accessible surface, and $N_{e,p}$ is the total number of electrons in the protein molecule and V_p is its partial molecular volume. Fig. 6 shows that the full profile $I_{pms}(K)$ is dominated by the scattering from the protein molecule with the effect of excluded solvent scattering included, ΔI_{pp}, and contribution from the 'h' region has a non-negligible magnitude. As described in the previous section, except at very small K, ΔI_{pp} is higher than I_{pms}, i.e. $\Delta I_{pp} \geq I_{pms}$ in the small K region. The primary source of the difference between them is the negative contribution from ΔI_{ph}. On the other hand, in the medium to large K region, the positive contribution of ΔI_{ph} and the negative contribution of ΔI_{hh} partially compensate with each other. Since the absolute value of ΔI_{ph} is larger than that of ΔI_{hh}, the magnitude relation is reversed as $\Delta I_{pp} \leq I_{pms}$. In contrast to those above, the contributions in which water in the 'b' region is involved are negligibly small in almost all K regions. This is the reason why the approximation $I_{pms} \approx \Delta I_{pp} + \Delta I_{ph} + I_{hh}$ holds very well. It will probably be due to a small error in the estimate of V_p that ΔI_{pb} has a significant contribution exceptionally in the very small K region.

To see the two hydration contributions, ΔI_{ph} and ΔI_{hh}, in more detail, the scattering profile for a model solute has been calculated. The solute particle is assumed a sphere having the same values of V_p and $N_{e,p}$ as those of myoglobin and a uniform electron density of $\bar{\rho}_p = N_{e,p}/V_p$. It is further assumed that there is a cavity layer with thickness 0.14 nm around the particle and, outside of it, a hydration layer with thickness 0.08 nm. The outside of the hydration layer is assumed a continuum

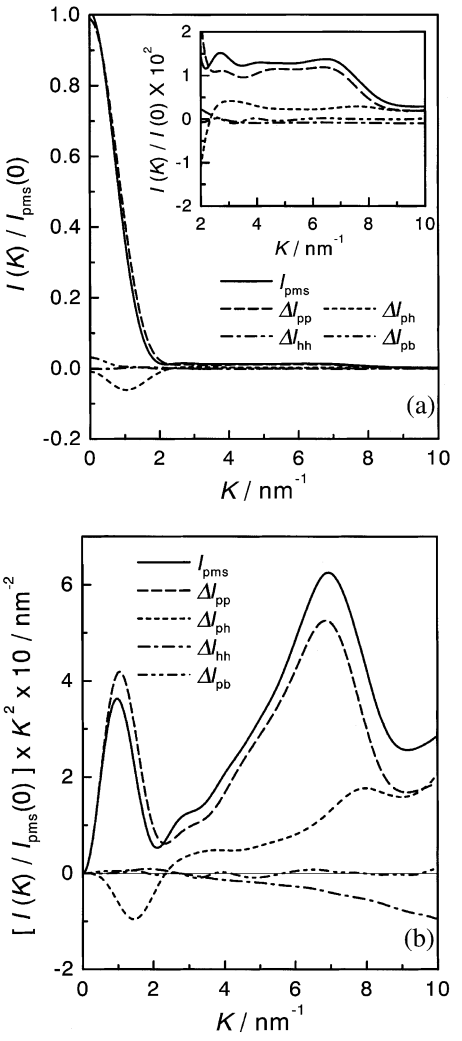

Fig. 6. Decomposition of the PMS profile of myoglobin into the six scattering components of ΔI_{pp}, ΔI_{ph}, ΔI_{hh}, ΔI_{pb}, ΔI_{hb} and ΔI_{bb}. (a) Linear plot and (b) Kratky plot. Both curves for ΔI_{hb} and ΔI_{bb} are not drawn, as they practically coincide with the horizontal line of $I(K)/I_{pms}(0) = 0$ in this scale of the ordinate.

with uniform electron density $\bar{\rho}$. The electron density in the hydration layer is assumed to be 2.56 times $\bar{\rho}$, under the condition of which the total number of excess electrons in the hydration layer and that of electrons excluded from the cavity layer just cancel with each other. Due to the assumptions of uniform electron density and the simple spherical shape of the solute, agreement of calculated PMS profiles for the model and myoglobin cannot be expected at large K. However, if the hydration contributions for the model solute reproduce well the corresponding ΔI_{ph} and ΔI_{hh} for myoglobin at small K, it will verify the importance of taking into account contributions from the cavity and the hydration layers for evaluating I_{pms} accurately.

Comparison of the hydration contributions for myoglobin and the model solute is made in Fig. 7. We can see from Fig. 7a, that both ΔI_{ph} and $\Delta I_{ph,\ model}$ have a negative peak at $K \approx 1.0\ nm^{-1}$ and are nearly constant at $K \geq 2.0\ nm^{-1}$. As seen from Fig. 7b, the peak of $\Delta I_{hh,\ model}$ at the smallest K side nearly coincides with that of ΔI_{hh} in its position and height. The oscillating behavior of $\Delta I_{hh,\ model}$ is a characteristic of the SXS profile of spherical particles. The reason why it is not observed in ΔI_{hh} will be that a myoglobin molecule is not of a spherical shape and has an uneven surface. Thus, both hydration contributions of the model solute reproduce qualitatively well those of myoglobin especially at small K. This result clearly demonstrates that the profiles of ΔI_{ph} and ΔI_{hh} in the small K region are affected strongly by the combined effect that there is a cavity layer around the solute protein and the hydration water outside the cavity has an electron density significantly higher than that of bulk water.

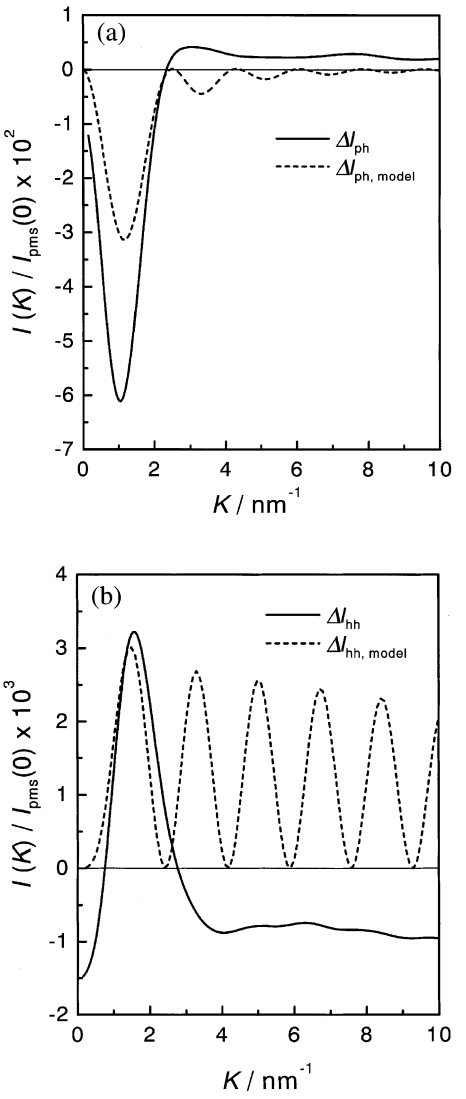

Fig. 7. Comparison of the hydration contributions of myoglobin and a spherical model solute. (a) Contributions from the interference effect between the solute and hydration water, ΔI_{ph} and $\Delta I_{ph,\ model}$, (b) contributions from the interference effect within the hydration region, ΔI_{hh} and $\Delta I_{hh,\ model}$.

3.4. Estimation of the thickness of the hydration layer

As stated in Section 3.2, if a hydration layer with an appropriate thickness is assumed, the PMS profile $I_{pms}(K)$ can be predicted with high accuracy by $\Delta I_{pp} + \Delta I_{ph} + \Delta I_{hh}$, in which only the solute protein and hydration water are taken into account explicitly. To estimate the actual thickness of the hydration layer, the approximate scattering profile, $\Delta I_{pp} + \Delta I_{ph} + \Delta I_{hh}$, of myoglobin and the squared sum of its deviation from the full profile I_{pms} are shown in Fig. 8a,b, where the thickness of the 'h' region is varied at intervals of 0.05 nm. In the figure, the distance d between the hb boundary and the accessible surface is used as a parameter for characterizing the thickness of the hydration

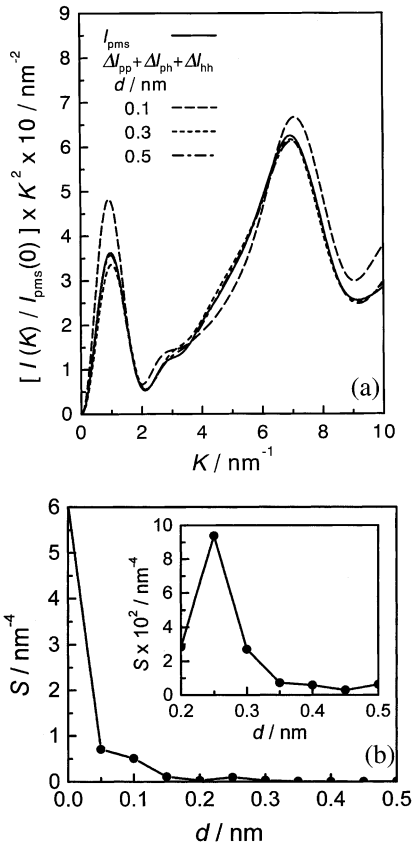

Fig. 8. Dependence of the approximate profile $\Delta I_{pp}+\Delta I_{ph}+\Delta I_{hh}$ for the full PMS profile $I_{pms}(K)$ on the assumed position of the hb boundary. The variable d is the distance between the accessible surface and the candidate for the hb boundary. (a) Kratky plot and (b) the squared sum of the deviations of $\Delta I_{pp}+\Delta I_{ph}+\Delta I_{hh}$ from $I_{pms}(K)$, $S=\sum\left[(\Delta I_{pp}+\Delta I_{ph}+\Delta I_{hh}-\Delta I_{pms})\times K^2/I_{pms}(0)\right]^2$, in the K range of $0\sim10$ nm^{-1}.

layer. It can be seen from Fig. 8 that the difference between the approximate and the full PMS profiles becomes negligibly small at $d\approx0.40$ nm or $t\approx0.54$ nm. This result indicates that, if only the hydration region with a thickness of two water layers is considered explicitly in the calculation of $\Delta I_{pp}+\Delta I_{ph}+\Delta I_{hh}$, the PMS profile can be predicted with high accuracy.

To see the above results in more detail, the space around the protein molecule was divided into shell-like regions at intervals of 0.02 nm. The average number of water molecules in each shell was determined using the result of MD simulation. Dividing the average number by the shell volume yields the average density of water as a function of the distance from the accessible surface of the protein, d. The function will be conveniently called the radial distribution function, RDF, though it has a meaning slightly different from the usual one. The RDF of water around the myoglobin molecule obtained from the MD simulation is shown in Fig. 9. Similar to that of pure water, the RDF of water for myoglobin exhibits a large maximum approximately 0.02 nm outside the accessible surface. It shows the first minimum and the second maximum at $d=0.17$ and 0.30 nm, respectively. The density of water is nearly equal to that of bulk water in the shell with d larger than 0.40 nm. Naturally, this agrees well with the result of analysis of the scattering profile. Thus, the thickness of hydration layer can be estimated as approximately that of two water layers, on average. It can also be seen that both the depth of the first minimum and the height of the second maximum are larger than those for pure water. This may indicate that water molecules near the surface of a protein have a

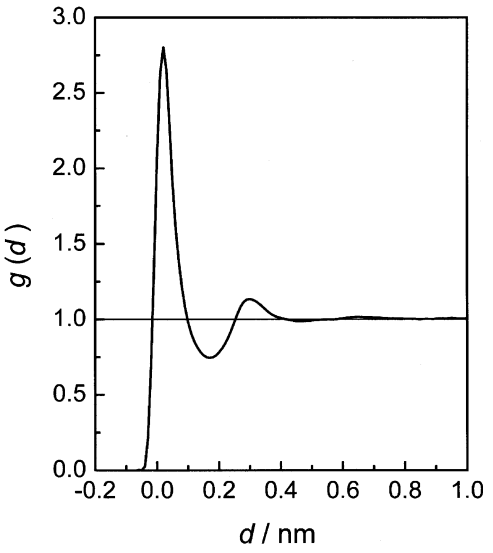

Fig. 9. Radial distribution function, $g(d)=\rho(d)/\bar{\rho}$, of water around a solute myoglobin in aqueous solution, where $\rho(d)$ and $\bar{\rho}$ are the density of water at distance d from the accessible surface and the average density in the bulk water region, respectively.

tendency to be located more regularly in the direction of the normal of the molecular surface.

3.5. Approximate incorporation of hydration effects

Dependence on K of the scattered X-ray intensity is determined definitely by the spatial distribution of electrons in solution. The SXS profile is given by the K dependence of intensity averaged over the spatial distributions of electrons. The average must be taken, not on the structure factor $F(K)$, but on the scattering intensity, $I_u(K)$ and $I_v(K)$. In most of the conventional methods of analysis, however, the effect of hydration water on the scattering profile is taken into account assuming a pre-averaged distribution of electron density for $F(K)$. In this section, accuracy of the approximate method is examined for three hydration models:

1. Effective continuum (EC) model. This is the most primitive model in which the outside of the effective molecular surface is approximated by a continuous medium with uniform electron density. The scattering profile predicted from this model is just the same as that of the I_{pp} approximation described in Section 3.2.
2. Radial distribution (RD) model. As done in the previous section, the space around the solute protein molecule is divided into shell-like regions and the average electron density of each shell is determined. Each shell is assumed to be filled with electrons having a constant density equal to the average density determined above. In this model, variation of the electron density in the radial direction is taken into account, but it is averaged out within each shell.
3. Spatial distribution (SD) model. When the coordinates of water molecules are sampled every t_s from an MD simulation of T_s, then $N_s(=T_s/t_s)$ sets of coordinate data are obtained. As $t_s=1$ ps and $T_s=500$ ps in our case, 500 coordinate data are generated for each water molecule. The whole set of coordinate data on the water molecules yields an approximate equilibrium distribution of water around the protein. In this model, the incident X-ray is assumed scattered from a continuous medium having the electron density distribution determined by the above distribution function of water. It is well known that a water molecule near to a charged or a neutral polar group is hydrogen-bonded to the group to show high occupancy at the position. As a result of it, the site has a high electron density, which results in a non-uniform distribution of electron density to contribute to X-ray scattering. The spatial distribution of hydration water specific to the solute is considered most accurately in this approximation.

Table 2
Estimates of the mean square radius R_{sq} of myoglobin from the full and three approximate PMS profiles

Hydration model	R_{sq}/nm
Full	1.75
EC: effective continuum	1.59
RD: radial distribution	1.77
SD: spatial distribution	1.74

The forward scattering intensity of a dilute solution is directly related to the partial molecular volume of the solute molecule as shown in Eq. (21). Hence, in order for all the hydration models to predict equal forward scattering intensity, they must yield an equal partial molecular volume. It is evident from the method of constructing the model that the RD model and the SD model yield the same partial molecular volume. As for the EC model the above condition is satisfied if the effective molecular surface is taken as the boundary between the 'p' region and the outside continuum. Values of R_{sq} for myoglobin obtained from Guinier analysis of the full PMS profile and the approximate profiles for the three hydration models in the range of $0 \le K^2 \le 0.4$ nm^{-2} are listed in Table 2. The SD and the RD models predict a value of R_{sq} nearly equal to and slightly larger than the exact value, respectively, while the EC model predicts as much as a 0.16-nm smaller value. This will be due to the situation that the effect of the non-uniform electron distribution around the protein molecule is partially considered in both SD and RD models, but it is not considered at all in the EC model. (The meaning of the word 'partially' will be clarified just below.) This result shows

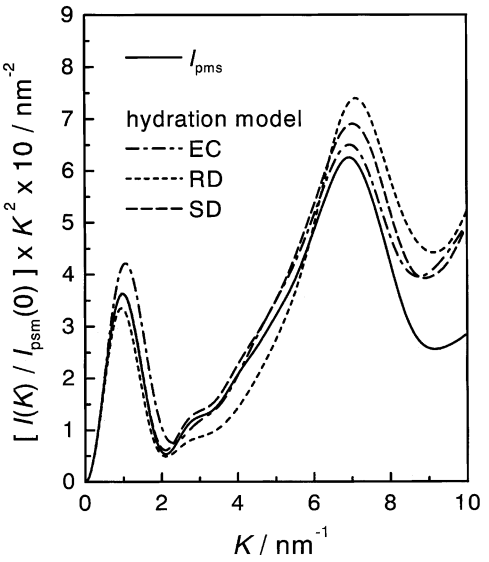

Fig. 10. Comparison of the full PMS profile and the approximate profiles predicted with three hydration structure models for myoglobin. Models: – – – –, full; – •– ••, EC (effective continuum); ·····, RD (radial distribution); – – – – , SD (spatial distribution).

that, to predict R_{sq} of a protein accurately, it is essential to choose a hydration model that takes into account the true characteristics of hydration structure as well as the partial molecular volume.

To see applicability of the hydration models at larger K values, the full and the approximate profiles are compared by Kratky plots in Fig. 10. We can see from Fig. 10 that, in the small K region, the SD model reproduces well the full PMS profile, while the EC and the RD models give a higher and a lower estimate for the scattering profile, respectively. It is already known from Fig. 5 that correct incorporation of the effect of hydration water is essential to obtaining a good approximation in the small K region. The above result shows that both EC and RD models are imperfect as the hydration structure model to predict the scattering profile in the small K region quantitatively. In the medium K region, the scattering profile is slightly overestimated by the SD and the EC models, while it is significantly underestimated by the RD model. The SD model, on the whole, is the best among the three as long as the small and medium K regions are concerned. However, even the prediction from the SD model deviates significantly from the full profile and the deviation increases with increase in K. In fact, all the models yield a remarkably high estimate in the large K region. We will discuss here the possible origin of this discrepancy. When the center of a water molecule is at a point in space, the probability that the center of another water molecule is found within a distance of 0.28 nm from the point is very low. The effect of such spatial correlation between water molecules is not considered at all in the EC model and only partially considered even in the SD model. This is a defect common to all of these approximations. It is obvious that the large error in the large K region results from this common defect in the three models. These results show that the pre-averaging approximation cannot fully incorporate the contribution from hydration water or the hydration effect. Thus, care should be taken for the accuracy of prediction when a pre-averaged hydration model is used to analyze experimental scattering profiles.

At the end of this section, we will comment on the program CRYSOL published in 1995 by Svergun et al. [29]. It is a program to evaluate PMS profiles of biological macromolecules from their atomic coordinates. Using a multipole expansion method, they succeeded in calculating PMS profiles very rapidly. In the following, discussion will be focused on its application to proteins. In CRYSOL, the scattering contribution from solvent molecules excluded by a protein molecule, which corresponds to $I_{\bar{p}\bar{p}}$ of our case, is considered by placing dummy solvent atoms at the positions of the protein atoms. The effect of hydration water, which corresponds to the hydration effect $\Delta I_{ph} + \Delta I_{hh}$ of our case, is approximately considered by assuming a hydration layer having an effective thickness of $\Delta = 0.3$ nm and a uniform electron density higher than that of bulk water by $\delta\rho$ around the molecular surface. An expansion factor for the dummy atoms and a density increment $\delta\rho$ are used to fit calculated profiles to experimental ones. As is apparent from this assumption, the hydration model taken in CRYSOL is one of the pre-averaged hydration models. Furthermore, the electron density distribution given by the model differs considerably

from that expected from our MD simulation. As is well known, the packing density of liquid water is significantly lower than that of a solid native protein, and the latter varies with protein species. As the dummy atoms are placed as described above, there must be cavities and overlaps between dummy atoms, which can never occur in pure water. As a result of it, calculation cannot reproduce the forward scattering intensity, and so the partial molecular volume of the solute protein correctly if the variable parameters are fixed constant. The two parameters are purely adjustable parameters, and their dependence on the molecular parameters is obscure. As shown in the RDF of Fig. 8, there must be a shell-like region near the protein surface in which the electrons of hydration water can never exist. It is essential to take into account correctly the contribution from this region for predicting accurately the mean square radius R_{sq} of the solute molecule. However, it is difficult to fully take into account the characteristics of hydration water by the model adopted in CRYSOL.

4. Conclusions

To examine how solvent water affects the scattering profile of a protein in aqueous solution, a theoretical framework was presented for evaluating it as the sum of interference effects of X-ray scattered from the three regions of: solute; hydration water; and bulk water. Molecular dynamic simulations were carried out on pure water and an aqueous solution of myoglobin to obtain 500 sets of data for the spatial distribution of water in both solvent and solution. With atomic coordinate data thus obtained, each of the X-ray scattering profiles of water and the aqueous protein solution was evaluated numerically as the sum of six different contributions. The partial molecular scattering (PMS) calculated for myoglobin was found to agree well with the experimental one. It was concluded from analysis of the calculated profiles that, to predict a PMS profile accurately, the following two are essential: (a) incorporating the contribution of excluded solvent scattering, that is, scattering from the solvent molecules excluded by introduction of a solute molecule, with the molecular nature and the spatial distribution of solvent molecules explicitly considered; and (b) incorporating the contribution of hydration water to scattering with its structural characteristics taken into account precisely. It was also found that considering only two molecular layers of water is enough to incorporate the hydration effect. Various approximate methods in which an averaged electron density distribution is assumed for evaluating not the scattering intensity, but the structure factor, have been used conventionally. However, care should be taken for using the approximation because calculated profiles deviate significantly from the true profile over the whole K region and, especially, the error may be larger at K larger than approximately 6 nm^{-1}.

Acknowledgments

The authors are grateful to Drs Yuzuru Hiragi, Kaoru Ichimura and Takashi Higurashi for their help in the SXS experiment of myoglobin. This work was supported by the Grants-in-Aid for Scientific Research on Priority Areas (A) (#11169216) and (C) (#12208006), and the Grant-in-Aid for Scientific Research (C) (#11680654) from the Ministry of Education, Culture, Sports, Science and Technology in Japan.

Appendix A: Method for generating a surface equidistant from the molecular surface

Let us denote the radius of the ith atom of protein as r_{0i}, which is given by Chothia et al. [37]. We will further suppose a pair of the hypothetical protein molecule having the same 3D structure as the real protein but different atomic radii and the corresponding probe sphere of variable radius r_p. Then, the surface composed of the points at distance t from the molecular surface of the real protein is given as follows:

1. When $0 \le t \le r_w$, it is defined by the molecular surface of the hypothetical protein consisting of the atoms with radius $r_i = r_{0i} + t$ for $i = 1 \sim N_{a,p}$ and the probe sphere of radius $r_p = r_w - t$.
2. When $t \ge r_w$, it is defined by the accessible surface of the hypothetical protein consisting of the atoms with radius $r_i = r_{0i} + t - r_w$ for $i = 1 \sim N_{a,p}$ and the probe water sphere of radius $r_p =$

r_w.

The number $N_{a,p}$ in (1) and (2) indicates the total number of atoms in the protein molecule.

References

[1] G.I. Makhatadze, P.L. Privalov, Energetics of protein structure, Adv. Protein Chem. 47 (1995) 307–425.

[2] G. Damaschun, H. Damaschun, K. Gast, C. Gernat, D. Zirwer, Acid denatured apo-cytochrome *c* is a random coil: evidence from small-angle X-ray scattering and dynamic light scattering, Biochim. Biophys. Acta 1078 (1991) 289–295.

[3] T.R. Sosnick, J. Trewhella, Denatured states of ribonuclease A have compact dimensions and residual secondary structure, Biochemistry 31 (1992) 8329–8335.

[4] M. Kataoka, Y. Hagihara, K. Mihara, Y. Goto, Molten globule of cytochrome *c* studied by small angle X-ray scattering, J. Mol. Biol. 229 (1993) 591–596.

[5] P.C. Calmettes, B. Roux, D. Durand, M. Desmadril, J.C. Smith, Configurational distribution of denatured phosphoglycerate kinase, J. Mol. Biol. 231 (1993) 840–848.

[6] G. Damaschun, H. Damaschun, K. Gast, et al., Cold denaturation-induced conformational changes in phosphoglycerate kinase from yeast, Biochemistry 32 (1993) 7739–7746.

[7] M. Kataoka, I. Nishii, T. Fujisawa, T. Ueki, F. Tokunaga, Y. Goto, Structural characterization of the molten globule and native states of apomyoglobin by solution X-ray scattering, J. Mol. Biol. 249 (1995) 215–228.

[8] S. Doniach, J. Bascle, T. Garel, H. Orland, Partially folded states of proteins: characterization by X-ray scattering, J. Mol. Biol. 254 (1995) 960–967.

[9] M. Kataoka, Y. Goto, X-Ray solution scattering studies of protein folding, Fold. Des. 1 (1996) R107–R114.

[10] M. Hoshino, Y. Hagihara, D. Hamada, M. Kataoka, Y. Goto, Trifluoroethanol-induced conformational transition of hen egg-white lysozyme studied by small-angle X-ray scattering, FEBS Lett. 416 (1997) 72–76.

[11] M. Kataoka, K. Kuwajima, F. Tokunaga, Y. Goto, Structural characterization of the molten globule of α-lactalbumin by solution X-ray scattering, Protein Sci. 6 (1997) 422–430.

[12] Y. Hagihara, M. Hoshino, D. Hamada, M. Kataoka, Y. Goto, Chain-like conformation of heat-denatured ribonuclease A and cytochrome *c* as evidenced by solution X-ray scattering, Fold. Des. 3 (1998) 195–201.

[13] Y.O. Kamatari, T. Konno, M. Kataoka, K. Akasaka, The methanol-induced transition and the expanded helical conformation in hen lysozyme, Protein Sci. 7 (1998) 681–688.

[14] D.J. Segel, A.L. Fink, K.O. Hodgson, S. Doniach, Protein denaturation: a small-angle X-ray scattering study of the ensemble of unfolded states of cytochrome *c*, Biochemistry 37 (1998) 12443–12451.

[15] Y.O. Kamatari, S. Ohji, T. Konno, Y. Seki, K. Soda, M. Kataoka, K. Akasaka, The compact and expanded denatured conformations of apomyoglobin in the methanol-water solvent, Protein Sci. 8 (1999) 873–882.

[16] T. Fujisawa, M. Kato, Y. Inoko, Structural characterization of lactate dehydrogenase dissociation under high pressure studied by synchrotron high-pressure small-angle X-ray scattering, Biochemistry 38 (1999) 6411–6418.

[17] L.A. Feigin, D.I. Svergun, Structure Analysis By Small Angle X-Ray and Neutron Scattering, Plenum Press, New York, 1987.

[18] O. Glatter, O. Kratky, Small Angle X-Ray Scattering, Academic Press, London, 1982.

[19] N. Niimura, Y. Minezaki, T. Nonaka, et al., Neutron Laue diffractometry with an imaging plate provides an effective data collection regime for neutron protein crystallography, Nat. Struct. Biol. 4 (1997) 909–914.

[20] M. Nakasako, Large-scale networks of hydration water molecules around bovine β-trypsin revealed by cryogenic X-ray crystal structure analysis, J. Mol. Biol. 289 (1999) 547–564.

[21] N. Matubayasi, L.H. Reed, R.M. Levy, Thermodynamics of the hydration shell. 1. Excess energy of a hydrophobic solute, J. Phys. Chem. 98 (1994) 10640–10649.

[22] L.R. Murphy, N. Matubayasi, V.A. Payne, R.M. Levy, Protein hydration and unfolding insights from experimental partial specific volumes and unfolded protein models, Fold. Des. 3 (1998) 105–118.

[23] J. Ninio, V. Luzatti, M. Yaniv, Comparative small-angle X-ray scattering studies on unacylated, acylated and cross-linked *Escherichia coli* transfer RNA_I^{Val}, J. Mol. Biol. 71 (1972) 217–229.

[24] B.A. Fedorov, O.B. Ptitsyn, L.A. Voronin, X-Ray diffuse scattering by proteins in solution. Consideration of solvent influence, J. Appl. Cryst. 7 (1974) 181–186.

[25] J.J. Mulle, Calculation of scattering curves for macromolecules in solution and comparison with results of methods using effective atomic scattering factors, J. Appl. Cryst. 16 (1983) 74–82.

[26] M.I. Pavlov, B.A. Fedorov, Improved technique for calculating X-ray scattering intensity of biopolymers in solution: evaluation of the form, volume and surface of a particle, Biopolymers 22 (1983) 1507–1522.

[27] K. Soda, Y. Miki, T. Nishizawa, Y. Seki, New method for incorporating solvent influence into the evaluation of X-ray scattering intensity of proteins in solution, Biophys. Chem. 65 (1997) 45–53.

[28] E.E. Lattman, Rapid calculation of the solution scattering profile from a macromolecule of known structure, Proteins Struct. Funct. Genet. 5 (1989) 149–155.

[29] D. Svergun, C. Barberato, M.H.J. Koch, CRYSOL — a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates, J. Appl. Cryst. 28 (1995) 768–773.

[30] T. Fujisawa, T. Uruga, Z. Yamaizumi, Y. Inoko, S. Nishimura, T. Ueki, The hydration of Ras p21 in

solution during GTP hydrolysis based on solution X-ray scattering profile, J. Biochem. (Tokyo) 115 (1994) 875–880.
[31] D.I. Svergun, S. Richard, M.H.J. Koch, Z. Sayers, S. Kuprin, G. Zaccai, Protein hydration in solution: experimental observation by X-ray and neutron scattering, Proc. Natl. Acad. Sci. USA 95 (1998) 2267–2272.
[32] D.T. Cromer, J.T. Waber, in: J.A. Ibers, W.C. Hamilton (Eds.), International Tables for X-Ray Crystallography, The Kynoch Press, Birmingham, 1974.
[33] A. Cheng, J.K.M. Merz, Application of the Nose–Hoover Chain Algorithm to the study of protein dynamics, J. Phys. Chem. 100 (1996) 1927–1937.
[34] W.D. Cornell, P. Cieplak, C.I. Bayly, et al., A second generation force field for the simulation of proteins, nucleic acids and organic molecules, J. Am. Chem. Soc. 117 (1995) 5179–5197.
[35] M.L. Connoly, Computation of molecular volume, J. Am. Chem. Soc. 107 (1985) 1118–1124.
[36] T.J. Richmond, Solvent accessible surface area and excluded volume in proteins, J. Mol. Biol. 178 (1984) 63–89.
[37] C. Chothia, Structural invariants in protein folding, Nature 254 (1975) 304–308.