ELSEVIER

Biophysical Chemistry 65 (1997) 45–53

Biophysical Chemistry

New method for incorporating solvent influence into the evaluation of X-ray scattering intensity of proteins in solution

Kunitsugu Soda *, Yōichiro Miki, Takayuki Nishizawa, Yasutaka Seki

Department of Bioengineering, Nagaoka University of Technology, Nagaoka, Niigata 940-21, Japan

Received 8 March 1996; revised 29 July 1996; accepted 6 September 1996

Abstract

A new method, the surface integration method, is presented for taking into account the influence of solvent on the intensity of X-ray scattered from proteins in solution. It requires no averaging numerically over the solute orientation. The solvent is modeled by a continuous medium with electrons of uniform density. This method is applied to amino acids, peptides and native proteins to confirm its effectiveness. The solvent influence on the normalized scattering intensity $I(K)/I(0)$ is more noticeable for larger solutes and at larger scattering angles, where $I(K)$ is the intensity of scattered X-ray with the magnitude of scattering vector $\boldsymbol{K}$.

Keywords: X-ray scattering; Solvent influence; Surface integration

1. Introduction

The solution X-ray scattering (SXS) method has long been used as an important means for determining global structural parameters of biopolymers in solution from the measurement of the angular dependence of X-ray intensity at small scattering angles [1,2]. Further, in recent years, with the prevalence of strong X-ray sources of synchrotron radiation, it is widely used to get information on the internal structure of proteins and nucleic acids from measurements at large angles. This ability of SXS has been successfully utilized to the characterization of protein structure in the nonnative state such as the unfolded and the molten globule states as well as that in the native state [3–6]. Thus, in addition to solution neutron scattering, SXS is expected to become important more and more as a tool for analyzing the solution structure of proteins.

The Guinier analysis of SXS data yields the mean squared radius R_{sq}, a quantity characterizing the spatial extension' of solute molecules. From differences of angular dependence of SXS for biopolymers, structural differences among these molecules with different primary structures or under different solution conditions have been investigated. Due to randomness of the spatial distribution of solute molecules in solution, it is impossible to determine their 3D structure from SXS data alone. Alternatively, in order to derive detailed structural information from measured data, it is necessary to calculate the angular dependence of scattering intensity for some possible structural models and to compare

* Corresponding author.

PII S0301-4622(96)02228-4

predictions from calculation with experimental data. Validity of each model is determined through this process and a refined model which reproduces experiment best is chosen. In this procedure, the influence of water must be correctly taken into account, because the solvent as well as the solute contributes to the X-ray scattering from an aqueous solution. Especially, in the calculation of scattering intensity at large angles, it is important to take into account the contribution of solvent as accurately as possible.

Generally, there are two levels for considering the solvent influence in the above calculation [1]: (1) The solvent is treated as a continuous medium with uniform electron density; (2) The solvation structure specific to the solute is explicitly considered. Actually, solvent molecules exhibit a nonuniform spatial distribution near the first hydration shell of the solute depending on the shape and polarity of its surface groups [7]. As a result, the R_{sq} value of a protein determined from Guinier analysis is in general slightly larger than the prediction from crystal structure data [8]. In spite of these facts, solvent water outside the first hydration shell has nearly the same uniform density as that of bulk water. It is also established that the difference between X-ray scattering intensities from a molecule in vacuum and that in solution can be mostly explained using the approximation that solvent water is a continuum having a uniform electron density. Hence, in this work, we discuss a method for taking into account the contribution of solvent in the continuum approximation: We propose a new approximate method of calculation based on an algorithm different from the conventional one, which is called the surface integration (SI) method. In the following, for practical importance, we are concerned only with proteins in aqueous solution. Extension to other molecules in nonaqueous solvent will readily be made.

Several methods have been developed for evaluating SXS intensity of biopolymers based on the continuum solvent model. In the effective atomic factor method [9], the solvent contribution is estimated by replacing each constituent atom of the solute by a sphere filled with electrons of constant density. Though it is simple enough to be easily applied, it has a serious disadvantage that the uniformity of electron density within the solute volume cannot be ensured: Density is multiplied in the region where more than two atomic spheres overlap, while the contribution from interatomic cavities is lost. This defect is removed in the 'cube method' developed by Ninio et al. [10] and Fedorov et al. [11], where the solute volume is divided into tiny cubes with uniform electron density. Refining the cube method, Muller [12] and Pavlov and Fedorov [13] developed the 'improved cube method' and 'the modified cube method', respectively, where the molecular volume defined as the total volume inaccessible to solvent is represented as an ensemble of cubic blocks. These methods enable taking into account contributions not only from isolated internal cavities but also from crevices of the solute where solvent molecules cannot enter due to their finite size. In any of these cube methods, however, the numerical volume integration over all cubes must be repeated for taking the average over many solute orientations. In the work of Pavlov and Fedorov [13], the average was taken with respect to 258 directions.

In this paper, we present a new approximate method where the solvent influence is evaluated through integrals over the solute surface exposed to solvent. Averaging over the solute orientation is made analytically, while the molecular volume is treated approximately.

2. Theory

We consider a system composed of a solute (u) and the solvent (v) surrounding it. The structure factor $F(\boldsymbol{K})$ of this system is given as

$$F(\boldsymbol{K}) = \int_u d\boldsymbol{r}\rho(\boldsymbol{r})e^{i\boldsymbol{K}\cdot\boldsymbol{r}} + \int_v d\boldsymbol{r}\rho(\boldsymbol{r})e^{i\boldsymbol{K}\cdot\boldsymbol{r}} \qquad (1)$$

where $\boldsymbol{K}$ is the scattering vector and $\rho(\boldsymbol{r})$ is the electron density at position $\boldsymbol{r}$ of a point P. Let the average electron density of the solvent be ρ_e and the atomic form factor of the constituent atom j at $\boldsymbol{r}_j$ be $f_j(K)$. In the usual approximation that every atom in a molecule has the same electron distribution as that of the isolated atom, Eq. (1) can be rewritten as

$$F(\boldsymbol{K}) = \sum_j f_j(K)e^{i\boldsymbol{K}\cdot\boldsymbol{r}_j} - \rho_e\int_u d\boldsymbol{r}e^{i\boldsymbol{K}\cdot\boldsymbol{r}} + \int_v d\boldsymbol{r}\{\rho(\boldsymbol{r}) - \rho_e\}e^{i\boldsymbol{K}\cdot\boldsymbol{r}} + \rho_e\int_{u+v} d\boldsymbol{r}e^{i\boldsymbol{K}\cdot\boldsymbol{r}} \qquad (2)$$

Numerical expressions of $f_j(\boldsymbol{K})$ for a number of atoms are given in the literature [14]. From the following relation

$$\int_{u+v} d\boldsymbol{r}\, e^{i\boldsymbol{K}\cdot\boldsymbol{r}} \underset{v\to\infty}{\to} (2\pi)^3 \delta(\boldsymbol{K}) \tag{3}$$

we can see that the fourth term of Eq. (2) has a finite contribution only to the forward scattering. The third term represents the contribution from the nonuniform electron distribution of water due to hydration of the solute, which is assumed to be negligible in this work as described in the previous section. Then, we have the following familiar expression

$$F(\boldsymbol{K}) \approx \sum_j f_j(\boldsymbol{K}) e^{i\boldsymbol{K}\cdot\boldsymbol{r}} - \rho_e \int_u d\boldsymbol{r}\, e^{i\boldsymbol{K}\cdot\boldsymbol{r}} \tag{4}$$

Let us introduce here a vector field $\boldsymbol{A}$ defined by

$$\boldsymbol{A} = -\frac{i\boldsymbol{K}}{K^2} e^{i\boldsymbol{K}\cdot\boldsymbol{r}} \tag{5}$$

Using the relation

$$\mathrm{div}\boldsymbol{A} = -\frac{i^2}{K^2}(\boldsymbol{K}\cdot\boldsymbol{K}) e^{i\boldsymbol{K}\cdot\boldsymbol{r}} = e^{i\boldsymbol{K}\cdot\boldsymbol{r}} \tag{6}$$

and applying the Gauss' theorem to $\boldsymbol{A}$, we obtain

$$\int_u d\boldsymbol{r} e^{i\boldsymbol{K}\boldsymbol{r}} = -\frac{i}{K^2} \int_S d\sigma K_n e^{i\boldsymbol{K}\cdot\boldsymbol{r}} \tag{7}$$

The right hand side of Eq. (7) is a surface integral over the solute surface, and K_n is the component of the vector $\boldsymbol{K}$ along the vector $\boldsymbol{n}$ at position $\boldsymbol{r}$ of a surface point, where $\boldsymbol{n}$ is the unit vector directed to the outward normal. From Eq. (4) and Eq. (7), we get

$$F(\boldsymbol{K}) = \sum_j f_j(K) e^{i\boldsymbol{K}\cdot\boldsymbol{r}_j} + \frac{i\rho_e}{K^2} \int_S d\sigma K_n e^{i\boldsymbol{K}\cdot\boldsymbol{r}} \tag{8}$$

Thus, the volume integral in Eq. (4) is replaced by the surface integral of K_n in Eq. (8). Using Eq. (8), the X-ray scattering intensity $I(\boldsymbol{K})$ averaged over the solute orientation is represented as

$$I(\boldsymbol{K}) = \left\langle |F(\boldsymbol{K})|^2 \right\rangle_\Omega = I_{uu}(K) + I_{uv}(K) + I_{uv}^*(K) + I_{vv}(K) \tag{9}$$

where

$$I_{uu}(K) = \sum_{j,k} f_j(K) f_k(K) \left\langle e^{i\boldsymbol{K}\cdot(\boldsymbol{r}_j - \boldsymbol{r}_k)} \right\rangle_\Omega \tag{10}$$

$$I_{uv}(K) = -\frac{i\rho_e}{K^2} \sum_j f_j(K) \int_S d\sigma \left\langle K_n e^{-i\boldsymbol{K}\cdot(\boldsymbol{r}-\boldsymbol{r}_j)} \right\rangle_\Omega \tag{11}$$

$$I_{vv}(K) = \frac{\rho_e^2}{K^4} \int_S d\sigma \int_S d\sigma' \left\langle K_n K_n' e^{i\boldsymbol{K}\cdot(\boldsymbol{r}-\boldsymbol{r}')} \right\rangle_\Omega \tag{12}$$

and $\langle\ \rangle_\Omega$ means the average with respect to the solute orientation.

2.1. $I_{uu}(K)$

We can easily show that $I_{uu}(K)$ is reduced to the well-known Debye formula describing the scattering intensity in vacuum:

$$I_{uu}(K) = \sum_{j,k} f_j(K) f_k(K) \frac{\sin(Kr_{jk})}{Kr_{jk}} \tag{13}$$

where r_{jk} is the distance between the centers of atom j at $\boldsymbol{r}_j$ and atom k at $\boldsymbol{r}_k$, that is,

$$\boldsymbol{r}_{jk} \equiv \boldsymbol{r}_j - \boldsymbol{r}_k,\ r_{jk} = |\boldsymbol{r}_{jk}| \tag{14}$$

2.2. $I_{uv}(K)$

We denote $\boldsymbol{n}$ as the unit vector in the direction of the outward normal at a point P_k of position $\boldsymbol{r}$ on the surface of atom k with radius a_k. The vector $\boldsymbol{\rho}$ directed from the center O_k of atom k to the point P_k is described as

$$\boldsymbol{\rho} = \boldsymbol{r} - \boldsymbol{r}_k = a_k \boldsymbol{n} \tag{15}$$

Then the vector $\boldsymbol{q}$ directed from the center O_j of atom j to P_k is expressed as

$$\boldsymbol{q} = \boldsymbol{r} - \boldsymbol{r}_j = \boldsymbol{\rho} - \boldsymbol{r}_{jk} = a_k \boldsymbol{n} - \boldsymbol{r}_{jk} \tag{16}$$

Considering the fact that averaging over solute orientations can be replaced by averaging over the directions of the variable scattering vector $\boldsymbol{K}$ with

the solute being fixed, we can rewrite Eq. (11) using elementary mathematics as

$$I_{uv}(K) = I^{*}_{uv}(K) = -\frac{1}{3}\rho_e \sum_j f_j(K) \sum_k \int_{S_k} d\sigma_k q_n j_{10}(K_q) \quad (17)$$

where q is the length of $\boldsymbol{q}$, and q_n is the component of $\boldsymbol{q}$ along $\boldsymbol{n}$. The function $j_{10}(x)$ is defined by

$$j_{10}(x) \equiv \frac{3(\sin x - x\cos x)}{x^3} \quad (18)$$

which is approximated at $x \leq 1$ as

$$j_{10}(x) \approx 1 - \frac{1}{10}x^2 + \frac{1}{280}x^4 - \cdots \quad (19)$$

From Eq. (16), we have

$$q^2 = r_{jk}^2 + a_k^2 - 2a_k(\boldsymbol{r}_{jk} \cdot \boldsymbol{n}) \quad (20)$$

$$q_n = a_k - (\boldsymbol{r}_{jk} \cdot \boldsymbol{n}) \quad (21)$$

Therefore, we obtain

$$q_n q^2 = a_k\left(r_{jk}^2 + a_k^2\right) - \left(r_{jk}^2 + 3a_k^2\right)(\boldsymbol{r}_{jk} \cdot \boldsymbol{n}) + 2a_k(\boldsymbol{r}_{jk} \cdot \boldsymbol{n})^2 \quad (22)$$

We will introduce here a function $j_{12}(x)$ defined by

$$j_{12}(x) \equiv \frac{10(1 - j_{10}(x))}{x^2} = \frac{10\{x^3 - 3(\sin x - x\cos x)\}}{x^5} \quad (23)$$

Using the following equality

$$\frac{1}{3}\sum_k \int_{S_k} d\sigma_k q_n = V_u \quad (24)$$

we can rewrite Eq. (17) as

$$I_{uv}(K) = -N_{e,u}\sum_j f_j(K) + \frac{1}{30}K^2\rho_e \sum_j f_j(K) \times \sum_k \int_{S_k} d\sigma_k q_n q^2 j_{12}(Kq) \quad (25)$$

where V_u is the molecular volume of the solute, and $N_{e,u}$, which is defined by

$$N_{e,u} = \rho_e V_u \quad (26)$$

is the total number of electrons within the van der Waals volume when it is filled with electrons of density ρ_e.

2.3. $I_{vv}(K)$

We will denote the points at $\boldsymbol{r}$ and $\boldsymbol{r}'$ on the surfaces of atom j and k as P_j and P_k, and the outward normal vectors of unit length at P_j and P_k as $\boldsymbol{n}$ and $\boldsymbol{n}'$, respectively. Respective vectors $\boldsymbol{\rho}$ and $\boldsymbol{\rho}'$ directed from the center O_j to P_j and from the center O_k to P_k are expressed as

$$\boldsymbol{\rho} = \boldsymbol{r} - \boldsymbol{r}_j = a_j\boldsymbol{n}, \quad \boldsymbol{\rho}' = \boldsymbol{r}' - \boldsymbol{r}_k = a_k\boldsymbol{n}' \quad (27)$$

The vector $\boldsymbol{s}$ directed from P_k to P_j is given by

$$\boldsymbol{s} = \boldsymbol{r} - \boldsymbol{r}' = \boldsymbol{r}_{jk} + \boldsymbol{\rho} - \boldsymbol{\rho}' = \boldsymbol{r}_{jk} + a_j\boldsymbol{n} - a_k\boldsymbol{n}' \quad (28)$$

Using Eq. (12) and elementary mathematics, we obtain the following equation

$$I_{vv}(K) = \frac{1}{15}\rho_e^2 \sum_{j,k} \int_{S_j} d\sigma_j \int_{S_k} d\sigma_k \times \left\{ \frac{5}{K^2}(\boldsymbol{n} \cdot \boldsymbol{n}')j_{10}(Ks) - (s_n \cdot s_{n'})j_{20}(Ks) \right\} \quad (29)$$

where s is the length of $\boldsymbol{s}$, and s_n and $s_{n'}$ are the components of $\boldsymbol{s}$ along $\boldsymbol{n}$ and $\boldsymbol{n}'$, respectively. The function $j_{20}(x)$ is defined by

$$j_{20}(x) \equiv \frac{15\{(3 - x^2)\sin x - 3x\cos x\}}{x^5} \quad (30)$$

which is approximated at $x \leq 1$ as

$$j_{20}(x) \approx 1 - \frac{1}{14}x^2 + \frac{1}{504}x^4 - \cdots \quad (31)$$

From Eq. (28), we have

$$s^2 = \left(r_{jk}^2 + a_j^2 + a_k^2\right) + 2a_j(\boldsymbol{r}_{jk} \cdot \boldsymbol{n}) - 2a_k(\boldsymbol{r}_{jk} \cdot \boldsymbol{n}') - 2a_j a_k(\boldsymbol{n} \cdot \boldsymbol{n}') \quad (32)$$

$$s_n = a_j + (\boldsymbol{r}_{jk} \cdot \boldsymbol{n}) - a_k(\boldsymbol{n} \cdot \boldsymbol{n}') \quad (33)$$

$$s'_n = a_k + (\boldsymbol{r}_{jk} \cdot \boldsymbol{n}') - a_j(\boldsymbol{n} \cdot \boldsymbol{n}') \quad (34)$$

Using the function $j_{12}(x)$ defined in Eq. (23) and taking into account the equality

$$\sum_j \int_{S_j} d\sigma_j \boldsymbol{n} = 0 \quad (35)$$

we can rewrite Eq. (29) as follows:

$$I_{vv}K = -\frac{1}{30}\rho_e^2 \sum_{j,k} \int_{S_j} d\sigma_j \int_{S_k} d\sigma_k \times \{s^2(\boldsymbol{n}\cdot\boldsymbol{n}')j_{12}(Ks) + 2(s_n \cdot s_n')j_{20}(Ks)\} \quad (36)$$

Further, we introduce two functions $j_{14}(x)$ and $j_{22}(x)$ defined by

$$j_{14}(x) \equiv \frac{28}{x^2}(1 - j_{12}(x)) = \frac{28}{x^7}\{x^5 - 10x^3 + 30(\sin x - x\cos x)\} \quad (37)$$

$$j_{22}(x) \equiv \frac{14}{x^2}(1 - j_{20}(x)) = \frac{14}{x^7}\left[x^5 - 15\{(3 - x^2)\sin x - 3x\cos x\}\right] \quad (38)$$

Using the easily verifiable relation

$$-\frac{1}{30}\sum_{j,k} \int_{S_j} d\sigma_j \int_{S_k} d\sigma_k \{s^2(\boldsymbol{n}\cdot\boldsymbol{n}') + 2(s_n \cdot s_n')\} = V_u^2 \quad (39)$$

we can rewrite Eq. (36) as

$$I_{vv}(K) = N_{e,u}^2 + \frac{1}{840}\rho_e^2 K^2 \sum_{j,k} \int_{S_j} d\sigma_j \int_{S_k} d\sigma_k s^4 \times (\boldsymbol{n}\cdot\boldsymbol{n}')j_{14}(Ks) + 4s^2(s_n \cdot s_n')j_{22}(Ks) \quad (40)$$

From those described above, we can estimate the scattering intensity $I(K)$ from Eq. (9) if we evaluate Eq. (13), Eq. (25) and Eq. (40). The first term of Eq. (25) does not depend on the spatial structure of the solute molecule, but depends only on its atomic composition. This term, together with the first term of Eq. (40), represents the primary effect of the difference in electron density between solute and solvent. It is the second term of each equation that reflects the effect of the 3D solute structure on the solvent influence. To evaluate each of these terms in $I_{uv}(K)$ and $I_{vv}(K)$, single and double surface integrals must be carried out over the surface of all atoms exposed to solvent. As it is generally difficult to execute these integrals analytically, we need to use some numerical method, which will be described in the next section.

3. Methods

In the following, we assume that the 3D structure of the solute composed of N nonhydrogen atoms is given by N sets of their atomic coordinates. Each hydrogen (H) atom is united with the nonhydrogen atom bonded to it. The form factor, and so, the radius of an atomic group consisting of a nonhydrogen atom and bonded H atom(s) is assumed to be the same as that of the isolated nonhydrogen atom with the number of electrons increased by the number of bonded H atoms. This expedient is taken because the radius of an atomic group must be consistent with its structure factor for correctly incorporating the contribution of solvent, but we have no means to determine the effective structure factor of such a group. This may be a poor approximation for the accurate estimation of the absolute SXS intensity, but will not change most of the qualitative conclusions described below.

Radii of C, O, N, S atoms and Fe^{3+} ion are taken as 0.16, 0.14, 0.15, 0.18 and 0.065 nm, respectively. The solvent is assumed to be water at 25°C, and the electron density of solvent ρ_e is $\rho_{w,25} = 333.29$ nm^{-3}. Structures of amino acids and the helical peptide of alanine decamer were generated using molecular structure parameters of AMBER force fields [15]. Atomic coordinate data of four proteins, S-peptide of RNase A, mellitin dimer, bovine pancreatic trypsin inhibitor (BPTI) and cytochrome c were taken from Protein Data Bank.

The gas-state contribution $I_{uu}(K)$ (Eq. (13)) can be evaluated straightforwardly through calculating Debye formula for $N(N+1)/2$ atomic pairs. To carry out surface integrals in $I_{uv}(K)$ (Eq. (25)) and $I_{vv}(K)$ (Eq. (40)) numerically, the spherical surface of each atom j ($j = 1 - N$) is divided into N_d segments with an equal area of $\delta\sigma_j = 4\pi a_j^2/N_d$, and a sampling point is taken at the center of gravity of each segment. The value of N_d was taken to be 400 in most of the following calculations. Surface inte-

grals for both equations are performed approximately by summing up all contributions from the sampling points exposed to solvent.

Respective values of q and s in the integrand of Eq. (25) and Eq. (40) vary with the sampling points. Therefore, in order to calculate, for example, $I_{vv}(K)$ with a high accuracy, we need to take a sufficiently large number of sampling points N_d and carry out the numerical integration for all pair of them on the exposed surface. Practically, $N_d = 400$ is a reasonable selection. The larger N_d value leads to the longer computation time as it is proportional to the square of N_d. Too large an N_d value would require enormously long time even for a medium-sized protein. Hence, the upper limit of the available computation time determines the maximum size of tractable proteins.

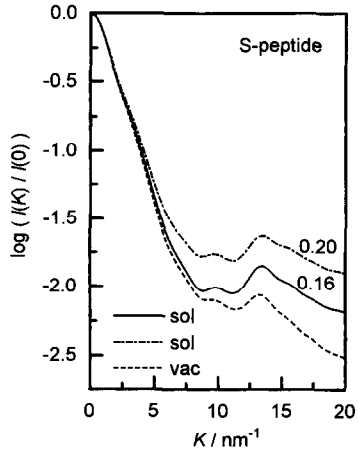

Fig. 2. Effect of the radius of C atom taken for calculation on the normalized intensity for S-peptide of RNase A. $N_d = 400$, and the radius indicated is in nm.

4. Results and discussion

To demonstrate the effectiveness of SI method described in the previous sections, we have performed numerical calculations for several molecules. Fig. 1 shows the K dependence of scattered X-ray intensity for an amino acid asparagine and an α-helical peptide of alanine decamer $(Ala)_{10}$: The normalized scattering intensity defined by the ratio $I(K)/I(0)$ is plotted against K. We can see that the normalized intensity is higher in solution than in vacuum over the whole K region, and the rate of its increase is larger at larger K. However, the characteristic feature of the K dependence observed in vacuum is kept almost unchanged, though the second peak for $(Ala)_{10}$ in solution shifts slightly to larger K than in vacuum. Three curves for asparagine with $N_d = 1600$, 400 and 72 coincide with each other within computational error. On the other hand, the normalized intensity of $(Ala)_{10}$ for $N_d = 72$ exhibits a slightly lower value than that for $N_d = 400$ which is practically equal to the exact value. This suggests that the artifact due to a smaller N_d may not be ignored for larger solutes and at larger K.

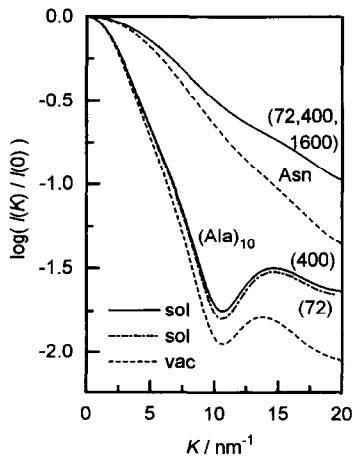

Fig. 1. K dependence of the normalized intensity for an amino acid, asparagine (Asn), and a peptide, alanine decamer $(Ala)_{10}$. The number in parenthesis represents the number of sampling points N_d taken for calculation, and 'sol' and 'vac' indicate solution and vacuum, respectively.

Dependence of the normalized intensity on the radius assumed for the carbon atom is shown in Fig. 2 for the S-peptide of RNase A, a 24-residue peptide having a partially helical structure. Due to the situation described in the previous section, 0.16 nm is considered to be a reasonable value for the C atom in our treatment. The radius of 0.20 nm yields significantly higher normalized intensity than that of 0.16 nm over the whole range of K studied. This will be due to the situation that too large a value assigned to the C atom increases the size of protein, which increases the volume of integration in Eq. (4), and the contribution of water is overestimated. This re-

sult indicates that, in order to estimate the scattering intensity for a solution with a high accuracy, values of atomic radius must be chosen carefully. Further, comparing Fig. 1 and Fig. 2, we can see that the larger and more complicated molecular structure leads to the more complex fine structure in the K dependence of scattering intensity.

Using two different plots for the K dependence of scattering intensity, three typical amino acids are compared in Fig. 3a and b. In the logarithmic plot of Fig. 3a, the dimensionless variable KR_{sq} is plotted as abscissa. All of the amino acids exhibit an oscillatory decreasing curve both in solution and vacuum. Naturally, in this plot, all the curves coincide with each other in the small K region. However, for every amino acid, there remains a significant difference between scattering profiles in solution and vacuum at large K region. Thus, the solvent effect considered here cannot be described as a simple effect on R_{sq}. Kratky plots for the amino acids are given in Fig. 3b. We can see that, while each curve for three amino acids in vacuum has a clearly recognized first peak, it practically disappears in solution only leaving a shoulder. On the other hand, the second peaks are mostly conserved. The first peak reflects the molecular size or the spatial extension of electron distribution of the molecule. Peaks at the larger K region originate from the short-range spatial correlation of electron density. This figure clearly shows that reduction of the difference in electron density between the molecular interior and the environment due to transfer from vacuum to solution decreases more greatly the scattering intensity at the smaller K region. Further, this indicates that care must be taken when we compare a Kratky plot predicted from calculation for vacuum with that obtained from experiment for solution.

Fig. 4 shows the change of normalized intensity with increasing the electron density of solvent ρ_e from that of water to the mean density of proteins. Such changes in electron density can be realized when we add organic liquids as alcohol or denaturants as urea and guanidine hydrochloride to aqueous solution. As seen from the figure, the normalized intensity increases with increase in ρ_e over the whole region of K, though the absolute intensity decreases (data not shown). The rate of increase is larger at larger K. General characteristics of the K dependence are kept unchanged with increase in ρ_e also in this case.

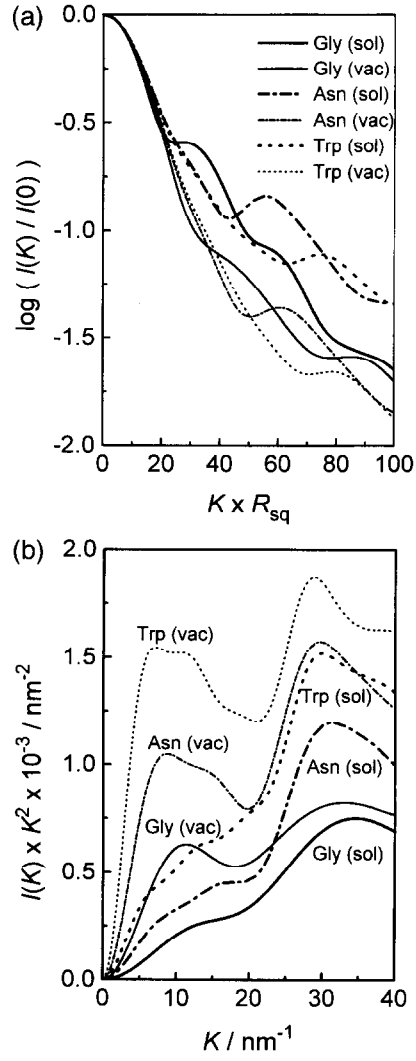

Fig. 3. K dependence of the normalized intensity for amino acids, glycine (Gly), asparagine (Asn) and tryptophan (Trp). $N_d = 400$: (a) Logarithmic plots, and (b) Kratky plots. Note that the abscissa of (a) is KR_{sq}.

Angular dependence of the normalized intensity is displayed in Fig. 5 for three proteins, mellitin dimer ($26 \times 2 = 52$ residues), BPTI (58 residues) and cytochrome c (103 residues). The difference between solution and vacuum for these proteins are significant in the smaller K region than those for amino acids and short peptides. Furthermore, the angular dependence exhibits more complicated fine struc-

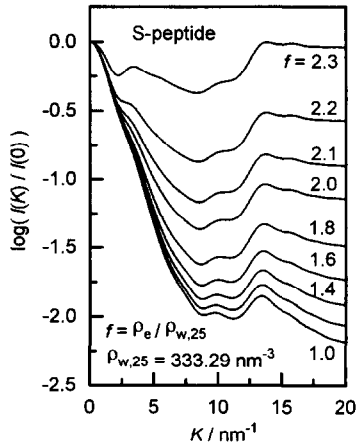

Fig. 4. Effect of the electron density ρ_e of solvent on the normalized intensity for S-peptide. The value of ρ_e is increased by steps of 20% the average density for water at 25°C, $\rho_{w,25}$ and $f = \rho_e / \rho_{w,25}$.

tures in the large K region. As a by-product of this work, it was also found that Eq. (24) and Eq. (39) are very useful for estimating the van der Waals volume of a protein, which is defined as the total volume occupied by all atoms of the protein.

Based on these results, physical basis of the solvent influence considered in this work will be discussed in the following: The X-ray scattering with scattering vector $\boldsymbol{K}$ originates from the $\boldsymbol{K}$ Fourier component of the electron density distribution. Two molecules of similar shape but different in size being

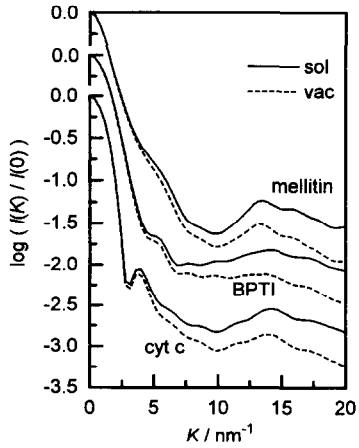

Fig. 5. K dependence of the normalized intensity for three proteins, mellitin dimer, BPTI and cytochrome c.

compared, the smaller one has in general smaller K components of electron density and yields the lower scattering intensity $I(K)$. On the other hand, the larger molecule has the higher proportion of contributions to the whole scattering from smaller K components. As a result of it, the normalized scattering intensity at K is generally lower for larger molecules as can be seen from the comparison of Asn and $(Ala)_{10}$ in Fig. 1. When a molecule in vacuum is dissolved into solvent, the difference in electron density between the molecule and its environment decreases, which results in decrease in scattering intensity over the whole region of K. The rate of this decrease is larger at smaller scattering angles or smaller K. Hence, as shown in figures, the normalized intensity is in general higher in solution than in vacuum. However, it is readily confirmed from displayed data that, compared at the same K, the ratio of the normalized intensity in solution to that in vacuum has a similar value nearly independent of molecular size. Thus, it is due to the difference in the ordinate scale of figures that differences between solution and vacuum for amino acids are seemingly larger than those for peptides and proteins.

The reason why the rate of decrease in scattering intensity due to introducing a molecule from vacuum into solution is larger at smaller K is as follows: As most of the electrons of each atom are localized around its nucleus, the real molecule has spatially nonuniform distribution of electron density. Each K component of electron distribution of the molecule contributes to the scattering at K. On the other hand, the X-ray scattering that contributes to the solvent effect considered in this work results from the scattering from a hypothetical body which has identical shape to the solute molecule and is filled with electrons of the same density as solvent. The distribution of electrons in the body is uniform and has no such irregularity as that in the real molecule. Therefore, with increasing K, the Fourier component of electron density distribution of the hypothetical body decreases more rapidly than that of the real molecule. Owing to this, the scattering from this body has the effect to cancel more strongly the scattering from the molecule at smaller K.

From those shown above, we conclude as follows: (1) The newly developed surface integration method has been confirmed effective to evaluate the X-ray

intensity scattered from proteins in solution without averaging numerically over the solute orientation. (2) There are significant differences between K dependencies of X-ray scattering from the solute in solution and that in vacuum, and the differences are larger for lager molecules and at larger K.

In spite of the effectiveness of our SI method, there remain several problems to be solved as follows: (1) In our treatment, contributions from internal cavities are included, but those from molecular crevices inaccessible to solvent are not taken into account. It is necessary to include correctly the contribution from these inaccessible regions, because most protein molecules are known to have small concave regions and clefts on their surface. (2) Developing some means to reduce the computation time is needed to apply this method to larger proteins with more than two hundred amino acid residues. The major factor responsible for the long computation time is the double surface integrals in Eq. (40). Development of an efficient algorithm to evaluate these integrals rapidly is urgently desired.

Related to the problem (1) above, it must be noted that general conclusions obtained in this work will not be qualitatively changed even if the contribution from the inaccessible region is accurately considered: If we take into account that solvent electrons cannot enter the region, it will lead to further increase in the volume with negative scattering amplitude in solution than our method. As a result of it, the tail of scattering curves should move not downward but upward further than in our case, which is similar to the result caused by an increase in the assumed atomic radius as shown in Fig. 2.

Acknowledgements

Authors express their thanks to Dr. T. Fujisawa for his valuable comments, and Dr. T. Nonaka for providing us with basic X-ray data. We are also grateful to E. Chikayama and K. Itoh for their helps in computer programming. This work was partly supported by the Grant in Aid (Nos. 06680646 and 06304051) for Scientific Research from the Ministry of Education, Science, Culture and Sport of Japan.

References

[1] O. Glatter and O. Kratky, Small Angle X-ray Scattering (Academic Press, London, 1982).
[2] J.A. Feigin and D.I. Svergun, Structure Analysis by Small-Angle X-ray and Neutron Scattering (Plenum, New York 1987).
[3] T. Fujisawa, T. Uruga, Z. Yamaizumi, Y. Inoko, S. Nishimura and T. Ueki, J. Biochem. (Jpn.) 115 (1994) 875.
[4] M. Kataoka, Y. Hagihara, K. Mihara and Y. Goto, J. Mol. Biol. 229 (1993) 591.
[5] M. Kataoka, I. Nishii, T. Fujisawa, T. Ueki, F. Tokunaga and Y. Goto, J. Mol. Biol. 249 (1995) 215.
[6] T. Konno, M. Kataoka, Y. Kamatari, K. Kanaori, A. Nosaka and K. Akasaka, J. Mol. Biol. 251 (1995) 95.
[7] E.N. Baker, in R.B. Gregory (Editor), Protein-Solvent Interactions (Marcel Dekker, New York, 1995).
[8] M. Kataoka, Private communication.
[9] R. Langridge, D.A. Marvin, W.E. Seeds, H.R. Wilson, C.W. Hooper, M.H.F. Wilkins and L.D. Hamilton, J. Mol. Biol. 2 (1960) 38.
[10] J. Ninio, V. Luzatti and M. Yaniv, J. Mol. Biol. 71 (1972) 217.
[11] B.A. Fedorov, O.B. Ptitsyn and L.A. Voronin, J. Appl. Cryst. 7 (1974) 181.
[12] J.J. Müller, J. Appl. Cryst. 16 (1983) 74.
[13] M.Yu. Pavlov and B.A. Fedorov, Biopolymers 22 (1983) 1507.
[14] D.T. Cromer and J.T. Waber, in J.A. Ibers and W.C. Hamilton (Editor), International Tables for X-ray Crystallography (The Kynoch Press, Birmingham, 1974).
[15] S.J. Weiner, P.A. Kollman, D.A. Case, U.C. Singh, C. Ghio, G. Alagona, S. Profeta, Jr. and P. Weiner, J. Am. Chem. Soc. 106 (1984) 765.