Improved Technique for Calculating X-Ray Scattering Intensity of Biopolymers in Solution: Evaluation of the Form, Volume, and Surface of a Particle

M. YU. PAVLOV, Institute of Protein Research, Academy of Sciences of the USSR, 142292 Poustchino, Moscow Region, USSR; and B. A. FEDOROV, Leningrad Mining Institute, Physics Department, 199026 Leningrad, USSR

Synopsis

An improved cube method has been developed for calculating the intensity of diffuse x-ray scattering of macromolecules in solution using a certain set of their atomic coordinates. The technique is based on the ideas of B. Lee and F. M. Richards [(1971) J. Mol. Biol. 55, 374-400] and Richards [(1977) Annu. Rev. Biophys. Bioeng. 6, 151-176] on the possibility of estimating the molecular and accessible surface of a particle by "rolling" a sphere, simulating a water molecule, on its molecular surface. It is shown that this technique is more advantageous than earlier versions of the cube methods. The improved technique for calculating scattering curves was utilized for several globular proteins, and for the first time, reliable scattering curves were obtained for protein-"bound" water complexes. In the case of globular proteins and tRNA, this technique has permitted a strict evaluation of their accessible surfaces, their volumes, and, apparently for the first time, their complete molecular surfaces.

INTRODUCTION

The technique of large-angle x-ray scattering developed for studying conformational rearrangements of biopolymers in solution1-4 is based on a calculation of the scattering intensity of a macromolecule in a solvent environment with nonzero electron density. This calculation is made from its atomic coordinates known from x-ray analysis. As has been shown experimentally by Stuhrmann5 and Ibel and Stuhrmann6 the solvent electron density significantly influences the scattering-curve profile, especially in the region of middle and large angles, i.e., at \( \mu \geq 0.2 \, \text{Å}^{-1} \), where \( \mu = (2\pi/\lambda) \sin \theta/2 \) (\( \lambda \) is the wavelength and \( \theta \) is the scattering angle). For a particle consisting of \( N \) atoms with the scattering factors \( f_j(\mu) \) and coordinates \( \vec{r}_j \), the scattering curve can be calculated from the equation

\[
I(\mu) = \langle I(\vec{\mu}) \rangle = \left\langle \sum_{j=1}^{N} f_j(\mu) e^{i\vec{r}_j} - \rho_s \phi(\vec{\mu}) \right\rangle^2
\]

where the averaging is done by all the possible orientations of the particle relative to the primary beam, \( \rho_s \) is the electron density of the solvent.
and

$$\phi(\mu) = \int_V e^{i\mu \mathbf{r}} \, d\mathbf{r}$$  \hspace{1cm} (2)

is the amplitude of scattering from the macromolecule volume inaccessible for solvent molecules. The main difficulty in calculating the scattering curve of macromolecules in solution is the estimation of this amplitude, and different techniques for calculating diffuse intensity vary largely in the method of determining the volume \(V\) by which integration is accomplished according to formula (2).

The most correct description of the volume, occupied by the particle in the solvent, can be obtained by the “cube method,” \(^7-9\) which allows filling all the places in the particle inaccessible to the solvent with rather small, densely packed cubes. The strictest approach proposed up to the present for such a description is the “modified cube method.” \(^9\) This technique uses cubes with an edge length of 1.3 Å and is evidently the most suitable for compact particles, especially for particles with a low surface-to-volume ratio when the inaccuracy of the surface description does not essentially influence the \(\phi(\mu)\) value. Examples of such particles are globular proteins.

On the other hand, the “modified cube method” (and more so, all the earlier techniques) proves to be insufficient quantitatively for at least two important types of structures: nucleic acids, which are comparatively “open” structures and for which the surface/volume ratio is markedly higher than for globular proteins of the same molecular weight; and complexes of globular proteins with bound water molecules, whose presence (in some cases in essential amounts) was revealed by x-ray and neutron structural analyses. This bound water, rather randomly located on the protein surface, leads to a considerable complication of the particle’s surface topography. Thus, it is necessary to develop a technique to give a much stricter description of the molecular surface of particles. It is evident that the “cube method” should be the basis of the technique, since only cubes give the dense packing and, consequently, homogeneous density within the particle volume that are necessary for calculating the integral (2). To solve this problem, it was first necessary to decrease significantly the size of a cube and, second, to consider the fact that atoms of a macromolecule have different van der Waals radii. Neither the first nor the second points could be realized using the algorithm of the “modified cube method.”

Lee and Richards\(^10\) have developed a technique for calculating the accessible particle surface by rolling a ball, imitating a water molecule, on the van der Waals surface of a particle. Our new technique for describing the shape and volume of a particle being a cube method is, to a great extent, based on the ideas of Lee and Richards. Therefore, in addition to the solution of diffraction tasks (1) and (2), our technique can be successfully used for a strict quantitative estimation of the main geometrical parameters of a particle. It allows a calculation of the accessible surface of a particle with
CALCULATING X-RAY SCATTERING INTENSITY

the same strictness as the method of Lee and Richards. Moreover, the same degree of accuracy can be achieved in estimating the molecular surface and volume, which are not evaluated by the method of Lee and Richards.

DESCRIPTION OF THE TECHNIQUE

According to Lee and Richards, the accessible surface of a particle is a totality of centers of water molecules in conditions of close contact with the van der Waals surface of the particle (Fig. 1). The surface on which balls, imitating water molecules, are rolling in conditions of close contact with the van der Waals particle surfaces is called the molecular surface. The volume bounded by the molecular surface is the volume inaccessible to the solvent. It is this volume that should be determined.

The main aim of our approach is to give a description of the molecular surface of a particle with small cubes, to estimate the molecular volume bounded by this surface, and to calculate the scattering amplitude according to formula (2). We believe that cubes with an edge length of about 0.3 Å are sufficiently small for a correct description of the molecular and accessible surfaces of any real molecule.

Let us mentally circumscribe a "hydrated" sphere with a radius equal to the sum of the van der Waals radii of the atom and a water molecule around every nonhydrogen atom of a macromolecule. A system of such overlapping spheres forms the volume of a "hydrated" macromolecule whose surface is the accessible surface of the particle. Then let us imagine that this "hydrated" molecule is fully placed within a parallelepiped consisting of small cubes with an edge length of about 0.3 Å. Every cube can be numbered 1, 2, or 0, depending on its location either within the "hydrated" molecule (1), on its surface (2), or beyond it (0). Cubes numbered...
2 form the accessible surface of the particle. Consequently, if every such cube is surrounded by a sphere with a radius of a water molecule and the cubes within this sphere are transformed from state 1 to state 0, we can pass from the volume of a hydrated molecule to the actual molecule volume consisting of cubes that remained in state 1. Cubes located on the boundary of regions 0 and 1 form a molecular surface. Using the algorithm described in Ref. 12, we can calculate the scattering amplitude according to Eq. (2) for the cubes forming the volume $V$ of the particle. Furthermore, the approach allows one to estimate both the accessible and the molecular surfaces of the particle.

For a complete realization of the above approach, it is sufficient in the suggested algorithm to use simultaneously only 11 sequential (i.e., at a distance of a cube edge length from each other) sections of a macromolecule. This number of sections is necessary for a complete description of a water molecule with cubes when it is centered in the middle section.

Let us consider a separate cycle of the program operation starting from the moment when a new (upper) section of the molecule is involved in the analysis and the lower one has already been released from its “hydrate coating.” First, we determine all the “influence” atoms for the new section, i.e., atoms whose “hydrated” volumes at least somehow touch this new section forming circles of different radii on it. Then, proceeding from the arrangement of these circles, all the cubes of the section are numbered 0, 1, or 2, according to the rules stated above; the coordinates of the cubes marked by 2 and, consequently, composing the accessible surface of the particle in this section, are placed in a separate array and stored.

The next stage is connected with the 6th section, which is equidistant from the upper and lower ones. For this section, coordinates of its cubes 2 are taken from the computer memory. Every such cube is surrounded by a sphere with the radius of a water molecule (this sphere comprises 11 sections), and all the cubes within the sphere located at different sections are transformed into a 0 state. Thus, if the upper section is just beginning its release from the hydrate coating, the lower one has already completed this process after carrying out the above procedure. Thus, we can, by gathering all the cubes of the lower section that have remained in state 1, describe the actual section of the molecular volume. Then, the new upper section is adjoined and the cycle is repeated until all the sections of the volume occupied by the molecule in solution are considered.

As a result, all the cubes 1 will form the molecular volume of the particle, and the cubes located on the boundary with other cubes (0) will outline the molecular surface of the particle.

Figure 1 represents a section of a particle with two water molecules “rolling” on its surface (a) and illustrates this section and the water molecules with cubes (b).

It should be noted that though the number of cubes necessary for a description of a “medium size” protein is $10^6$ in the approach presented, the computer time required for their calculation only slightly exceeds that of
an analogous calculation with the “modified cube method,” for which the number of cubes is about $10^4$. This can be explained by the fact that the objects analyzed in the proposed technique are cubes of only the accessible surface, whereas the earlier method takes into consideration all the cubes that can be centers of water molecules.

**CALCULATION OF SCATTERING INTENSITY**

Knowing the coordinates of all cubes $1$, it is not difficult to calculate the amplitude of scattering from the molecular volume of the particle by Eq. (2).

To accelerate such a calculation, the cubes neighboring along one of the directions of the section (e.g., $Y$) were united into parallelepipeds with a length of $2b_j = 2an$. Then, the calculation was made using the equation in Ref. 12:

$$
\phi(\mu) = 8 \frac{\sin(\mu_x a) \sin(\mu_y a)}{\mu_x \mu_y \mu_z} \sum_j \sin(\mu_y b_j) e^{i(\mu \bar{r}_j)}
$$

where $\bar{r}_j$ are coordinates of the center of the $j$th parallelepiped, $2a$ is a cube edge, and $n$ is the number of cubes in a parallelepiped.

To obtain the scattering intensity in solution, the square of the scattering-amplitude module was averaged in reciprocal space by a sphere [more exactly, by a semisphere using $I(\mu) = I(-\mu)$] with the radius $|\mu|$. In practical calculations, the intensity was averaged by 129 points uniformly distributed on the surface of a semisphere, and it was shown that for proteins, a further increase of the number of points does not change the scattering curves in the considered range of scattering angles (up to $\mu = 0.7$ Å$^{-1}$).

An important test for strictness of the technique is a comparison of the scattering intensity (and also the volume and accessible and molecular surfaces), calculated for the same particle differently oriented relative to the system of coordinates in which the cubes are built. It was found that divergence between the calculated curves did not, as a rule, exceed 1%, and the divergence between the calculated volumes and accessible and molecular surfaces did not exceed 0.3, 0.8, and 1.4%, respectively. It should be noted that in the “modified cube method,” the corresponding changes of the curve and the volume were usually much more pronounced.

In actual calculations we used a cube edge-length of $2a = 0.279$ Å. A water molecule was described by a system of 651 cubes that approximated a sphere with a radius of 1.5 Å. This value agrees with the minimum distance from the center of the water molecule to its van der Waals surface. The values for the van der Waals radii of atoms and atom groups were taken from Bondi$^{13}$ (Table I) and are very close to those used by Richards.$^{14}$

A choice of the van der Waals radii of atoms affects, of course, the calculated profile of the scattering curve. However, we have shown that a minor alteration of these radii ($\pm 0.1$ Å) practically does not change the
intensity of scattering, provided the particle volume remains unchanged.

**SCATTERING INTENSITY OF GLOBULAR PROTEINS**

The new technique has been used for a calculation of scattering curves of a number of globular proteins and for a comparison of the scattering

---

**TABLE I**

<table>
<thead>
<tr>
<th>Type of Atom or Group</th>
<th>Bondi (Ref. 13)</th>
<th>Richards (Ref. 14)</th>
<th>Lee &amp; Richards (Ref. 10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;C=</td>
<td>1.74</td>
<td>1.70</td>
<td>1.80</td>
</tr>
<tr>
<td>&gt;CH</td>
<td>2.00</td>
<td>2.00</td>
<td>1.80</td>
</tr>
<tr>
<td>-CH₂</td>
<td>2.00</td>
<td>2.00</td>
<td>1.80</td>
</tr>
<tr>
<td>CH₃</td>
<td>2.00</td>
<td>2.00</td>
<td>1.80</td>
</tr>
<tr>
<td>-NH</td>
<td>1.65</td>
<td>1.70</td>
<td>1.55</td>
</tr>
<tr>
<td>NH₂</td>
<td>1.75</td>
<td>-</td>
<td>1.80</td>
</tr>
<tr>
<td>NH₃⁺</td>
<td>-</td>
<td>2.00</td>
<td>1.80</td>
</tr>
<tr>
<td>O=</td>
<td>1.50</td>
<td>1.40</td>
<td>1.52</td>
</tr>
<tr>
<td>OH</td>
<td>-</td>
<td>1.60</td>
<td>1.80</td>
</tr>
</tbody>
</table>

---

**Fig. 2.** Theoretical scattering curves of sperm whale myoglobin calculated by the “modified cube method” (---) and the improved technique (----).
Fig. 3. Theoretical scattering curves of hen egg-white lysozyme calculated by the "modified cube method" (---) and the improved technique (----) using protein coordinates in tetragonal (a) and triclinic (b) forms.

curves obtained with curves calculated by the "modified cube method." As has been expected, the divergence between the scattering curves calculated by the old and improved techniques is, in most cases, insignificant, although in some cases it becomes essential for a quantitative comparison of scattering curves with experimental data.

Figure 2 shows sperm whale myoglobin scattering curves calculated by the old and improved techniques. The curves differ slightly quantitatively,

Fig. 4. Theoretical scattering curves of hen egg-white lysozyme calculated using protein coordinates in tetragonal (----) and triclinic (-----) crystal forms by the improved technique (a) and the "modified cube method" (b).
though the difference does not influence the previously suggested structural interpretation of the divergence between experimental and theoretical scattering curves.\textsuperscript{1,4}

Figure 3 shows scattering curves calculated by the old and improved techniques using atomic coordinates of hen egg white lysozyme determined in tetragonal\textsuperscript{15} and triclinic\textsuperscript{16} crystal forms. For a tetragonal form [Fig. 3(a)], the curves calculated by both techniques are compatible, while they display noticeable quantitative differences for a triclinic form [Fig. 3(b)]. It is important to note that scattering curves calculated by the improved technique for both forms of lysozyme practically coincide up to $\mu = 0.55$ Å\(^{-1}\), diverging slightly only in the region of the right-hand maximum [Fig. 4(a)], whereas the old technique gives significant divergence between the curves beginning with $\mu = 0.25$ Å\(^{-1}\) [Fig. 4(b)]. Structural differences of lysozyme molecules in the triclinic and tetragonal forms are very small; therefore, the similarity of scattering curves of both forms is evidence for the validity of the improved technique.

**SCATTERING INTENSITY OF GLOBULAR PROTEINS WITH BOUND WATER MOLECULES**

A comparison of scattering curves calculated by the improved technique with and without consideration for the water molecules bound to protein is of special interest. Figures 5–7 show the respective scattering curves for ribonuclease S, hen egg white lysozyme, and carp parvalbumin B. For the first two proteins, the scattering curves are comparatively similar (with and without “bound” water), but for carp parvalbumin B, these curves display noticeable quantitative differences. The most likely explanation is a different relative contribution of the mass of water to the molecular mass of the protein–water complex: the $q = M_w/M_p$ value ($M_w$ is the mass of “bound” water, $M_p$ the molecular mass of protein) is 0.061, 0.122, and 0.216 for ribonuclease S, lysozyme, and carp parvalbumin B, respectively.

In these calculations, van der Waals radius of bound water molecules was taken as 1.625 Å proceeding from the radius of the hydroxyl group used in Ref. 14. At this value, the mean electron density of bound water is close to that of the water solvent. However, as shown for the example of carp parvalbumin B (Fig. 7), variation of the van der Waals radius of bound water does not really influence the large-angle region of the scattering curve.

It should be emphasized that reliable scattering curves for the protein-bound water complexes can be obtained only by use of the improved cube method which allows a detailed description of a jagged surface of these complexes. Thus, Fig. 8 represents scattering curves for the carp parvalbumin B–water complex, as compared with the scattering curve for the same complex calculated by the “modified cube method.”\textsuperscript{9} It can be seen that in this case the “modified cube method” is unsuitable for a calculation.
of the scattering curve for the complex, since it distorts the scattering curve even in the small-angle region.

**CALCULATION OF THE VOLUME AND MOLECULAR AND ACCESSIBLE SURFACES OF GLOBULAR PROTEINS AND tRNA**

The improved cube method can also be used for a quantitative estimation of the volume, and molecular and accessible surfaces of macromolecules in solution.

It follows from the above algorithm that the accessible surface of a particle can be calculated using a joint array of cubes 2, while the molecular surface can be calculated using the array of cubes 1 bordering on cubes 0. Knowing the number of cubes situated on the surface, it is easy to calculate the surface value. For this it is necessary to determine the conversion coefficient $\sigma$ in the formula

$$A_s (\text{Å}^2) = \sigma_A \cdot N_A$$
Fig. 6. Comparison of theoretical curves calculated by the improved technique for hen egg-white lysozyme (---) and the complex of lysozyme with 102 water molecules (---).

where $A_s (\AA^2)$ is the surface in $\AA^2$ and $N_A$ is the number of cubes on the surface. The accessible surface is composed of pieces of spheres with radii varying from 3 to 3.5 $\AA$, and the molecular surface from pieces of spheres with radii varying from 1.5 to 2 $\AA$. Therefore, for the accessible and molecular surface, it is expedient to choose the mean values of the coefficient $\sigma$ for spheres with radii from 3 to 3.5 $\AA$ and 1.5 to 2 $\AA$, respectively. The calculated value of $\sigma_A$ is 0.101 $\AA^2$/cube for the accessible surface ($A_s$) and $\sigma_M$ is 0.097 $\AA^2$/cube for the molecular surface ($M_s$).

A reliable test for the improved technique is a comparison of our calculations of the accessible surface with the results of Lee and Richards.\textsuperscript{10} For such a comparison one should use the values of van der Waals radii used by Lee and Richards, which are underestimated as compared with the data of Bondi\textsuperscript{13} (see Table I). Table II shows accessible surfaces $A_s$ for several proteins calculated from these radii. The magnitude of error in our method compared with that of Lee and Richards\textsuperscript{10} is less than 1%, which confirms its high validity.

Table II also includes accessible and molecular surfaces as well as volumes calculated from the van der Waals radii given by Bondi.\textsuperscript{13} The reliability of these radii is confirmed by the fact that the calculated protein volumes are close to the value obtained from the experimentally measured partial
### TABLE II
Geometric Characteristics of Proteins and tRNA<sub>Phe</sub> Calculated Using the Improved Technique

<table>
<thead>
<tr>
<th>Protein</th>
<th>Molecular Weight&lt;sup&gt;a&lt;/sup&gt;</th>
<th>van der Waals Radii Used by Lee and Richards (Ref. 10)</th>
<th>Volume Calculated from Experimental Partial Specific Volume, V&lt;sub&gt;b&lt;/sub&gt; (Å&lt;sup&gt;3&lt;/sup&gt;)</th>
<th>van der Waals Radii Used by Bondi (Ref. 13)</th>
<th>M&lt;sub&gt;b&lt;/sub&gt; (Å&lt;sup&gt;2&lt;/sup&gt;)</th>
<th>M&lt;sub&gt;d&lt;/sub&gt; (Å&lt;sup&gt;2&lt;/sup&gt;)</th>
<th>M&lt;sub&gt;d&lt;/sub&gt;/A&lt;sub&gt;d&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carp parvalbumin B (3CPV)</td>
<td>11,520</td>
<td>7010&lt;sup&gt;c&lt;/sup&gt; 6993 13,907</td>
<td>15,730&lt;sup&gt;11&lt;/sup&gt; 15,448 5086</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ribonuclease S (1RNS)</td>
<td>13,690</td>
<td>6710&lt;sup&gt;c&lt;/sup&gt; 6743 15,243</td>
<td>16,900&lt;sup&gt;17&lt;/sup&gt; 16,887 4940</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hen egg-white lysozyme (6LYS)</td>
<td>14,300</td>
<td>8020&lt;sup&gt;e&lt;/sup&gt; 8000 19,192</td>
<td>21,980&lt;sup&gt;17&lt;/sup&gt; 21,460 7721 5826</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sperm whale myoglobin (1MBN)</td>
<td>18,610</td>
<td>9091 19,960</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bacteriophage T4 lysozyme (1LZM)</td>
<td>21,630</td>
<td>13,347&lt;sup&gt;d&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adenylate kinase (2ADK)</td>
<td>23,300</td>
<td>13,347&lt;sup&gt;d&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Actinidin (1ACT)</td>
<td>23,425</td>
<td>13,347&lt;sup&gt;d&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Papain (8PAP)</td>
<td>25,900</td>
<td>13,347&lt;sup&gt;d&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tRNA&lt;sub&gt;Phe&lt;/sub&gt; (4TNA)</td>
<td>27,540</td>
<td>13,347&lt;sup&gt;d&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elastase (1EST)</td>
<td>34,560</td>
<td>13,347&lt;sup&gt;d&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subtilysin BPN (1SBT)</td>
<td>34,700</td>
<td>13,347&lt;sup&gt;d&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thermolysin (2TLN)</td>
<td>34,700</td>
<td>13,347&lt;sup&gt;d&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carboxypeptidase A (1CPA)&lt;sup&gt;f&lt;/sup&gt;</td>
<td>41,550&lt;sup&gt;24&lt;/sup&gt;</td>
<td>41,620 11,434 9327</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> The molecular weight is calculated from the composition.
<sup>b</sup> The volume is calculated from experimental specific volumes and the protein molecular weight using formula \( V = \bar{v}M_s/N_A \), where \( \bar{v} \) is a partial specific volume, \( M_s \) is a protein molecular weight, and \( N_A \) is Avogadro's number.
<sup>c</sup> Calculated by Alden and Kim (Ref. 25) from their own values for van der Waals radii.
<sup>d</sup> Calculated by Lee and Richards (Ref. 10).
<sup>e</sup> Calculated by Alden and Kim (Ref. 25) from their own values for van der Waals radii.
<sup>f</sup> Calculations are done for the complex of carboxypeptidase A with pseudo-substrate Gly-Tyr.
Fig. 7. Comparison of theoretical curves calculated by the improved technique for carp parvalbumin B (---) and the complex of parvalbumin with 138 water molecules for two different radii of fixed water: (---, \( R_w = 1.595 \) Å) and (--, \( R_w = 1.625 \) Å). In all cases, \( \alpha = 0.075 \) Å.

specific volumes (Table II). On the other hand, the protein volumes determined from the van der Waals radii used by Lee and Richards are noticeably lower (Table II).

Table II shows that the molecular and accessible surfaces for different proteins are not proportional to each other. The ratio \( M_a/A_s \) varies from 0.739 (ribonuclease S) to 0.816 (carboxypeptidase A). The mean value of this ratio for proteins (Table I) is 0.722, with a dispersion of 0.028. This ratio characterizing the degree of "jaggedness" of the particle surface may become an important parameter for a description of the geometrical properties of macromolecules.

Coordinates of the protein nonhydrogen atoms necessary for the calculations described in the present paper have been taken from the Protein Data Bank. In Table II the proteins names are accompanied by labels of corresponding sets of coordinates from the Protein Data Bank.

**DISCUSSION**

The improved technique has allowed us to make a strict calculation of scattering curves for complexes of globular proteins with bound water
molecules and, thus, to estimate the effect of bound water on the scattering intensity of proteins. However, there are grounds for assuming that the values of this influence, given in Fig. 6 and especially in Fig. 7, are overestimated. Indeed, according to nmr data, the protein surface has almost no fixed water, and the majority of molecules composing the first hydrate coating have a rather high mobility. It has also been shown that only about 30–40 water molecules of this coating of a medium-sized protein (molecular weight of about 20,000) have relaxation time comparable to the rotational time of the protein as a whole.

It is natural to assume that water molecules of the second hydrate coating are much more weakly bound to the protein than those of the first one. This allows one to eliminate water molecules that are not involved in the first hydrate coating. The number of such molecules in the case of carp paravalbumin B is 80 out of the total of 138 water molecules. Figure 9 shows the scattering curve of the complex of carp paravalbumin B with the remaining 58 water molecules of the first hydrate coating. As seen, taking into account only the first hydrate coating significantly decreases the discrepancy of the scattering curves for “hydrated” and “nonhydrated” protein molecules. Therefore, it can be supposed that the influence of bound water on the scattering intensity of the protein in solution is not great and that

---

**Fig. 8.** Theoretical scattering curves of the complex of carp parvalbumin with 138 water molecules calculated by the improved technique (—) and the "modified cube method" (---).
it is most expedient to evaluate this influence for separate proteins by considering only the first hydrate coating.

The possibility of a detailed description of the molecular surface and volume of a biopolymer is not only of interest for large-angle x-ray scattering. It is also known that protein tertiary structure and its alterations are largely determined by the interaction of the protein with the solvent and this interaction, in turn, significantly depends on the surface value of the protein molecule. Chothia\cite{30} and later Janin and Chothia\cite{30} attempted to connect a change in the protein molecule’s accessible surface with the free energy of the protein interaction with the solvent and, consequently, with a choice of the most probable conformations of the protein in solution. However, the molecular surface apparently describes the protein interaction with the solvent more adequately, since it takes into account (due to its concave patches) a special energetic state of the solvent molecules, which form two or more contact surfaces with the protein molecule (see Fig. 1). The method of molecular surface evaluation proposed in Ref. 31 permits estimation of the molecular surface of the contact site of two particles but is inefficient for a calculation of a complete molecular surface or its highly jagged parts. The development of a strict technique
for calculating the complete molecular surface and volume of a particle presented in this paper points to the possibility of using these geometric parameters to describe the interaction of biopolymers with a solvent.

The authors thank Professor O. B. Ptitsyn for valuable discussions and Drs. I. N. Serdyuk and Yu. N. Chirgadze for reading the manuscript and for helpful criticism.

References


Received March 8, 1982
Accepted November 17, 1982