Acta Physica Academiae Scientiarum Hungaricae, 53 (1—2), pp. 105—119 (1982)

THE INVESTIGATION OF THE STRUCTURE OF BIOMOLECULES BY MEANS OF SMALL-ANGLE SCATTERING*

B. K. VAINSHTEIN, L. A. FEIGIN and D. I. SVERGUN

INSTITUTE OF CRYSTALLOGRAPHY, ACADEMY OF SCIENCES OF THE USSR
MOSCOW, USSR

The general features of the diffraction techniques used for the investigations of biological macromolecules are considered. The possibilities of small-angle scattering as a method of structure determination are analysed. The new method of direct structure analysis of small-angle scattering data is described and an example of its practical application is presented.

The study of the structure of biological macromolecules, i.e. proteins, nucleic acids, nucleoproteins, polysaccharides and their associations into more complex systems such as viruses, ribosomes, chromosomes, membranes and so on represents a rather important problem in molecular biology. The physical methods to treat these problems are based on the scattering of X-rays, electrons and neutrons, respectively. After discussing the general principles of diffraction techniques, their possibilities and limitations, the paper discusses several methods of structure determination by means of small-angle scattering (SAS). Further on some data on the structure of a few biomolecules will be presented.

Our ultimate task aims at obtaining a pattern of the spatial structure $\rho(\mathbf{r})$ of the object under investigation. By means of optics as, for instance, in a transmission or electron microscope, one obtains directly an image of the object. The image formation can be presented according to Abbe's scheme as demonstrated in Fig. 1. However, depending on the nature of radiation used for analysing the structure under investigation the realization of the full optical pattern is sometimes impossible, though its first stage, i.e. the observation of scattered, diffracted radiation may be always accomplished. In diffraction experiments one first measures the intensities $I(\mathbf{s})$ or moduli of amplitudes $F(\mathbf{s})$ of the scattered radiation, subsequently the diffracted beam is collected with the aid of lenses to form an image of the object, or, to be more precise, to present an image of its projection $pr_{\mathbf{k}}\rho(\mathbf{r})$ along the vector $\mathbf{k}$ of the initial wave (whenever possible). In the diffraction structure analysis the transition from the intensities to the structure of an object $\rho(\mathbf{r})$ is not realized physically, but it may

* Dedicated to Prof. I. Tarján on his 70th birthday.

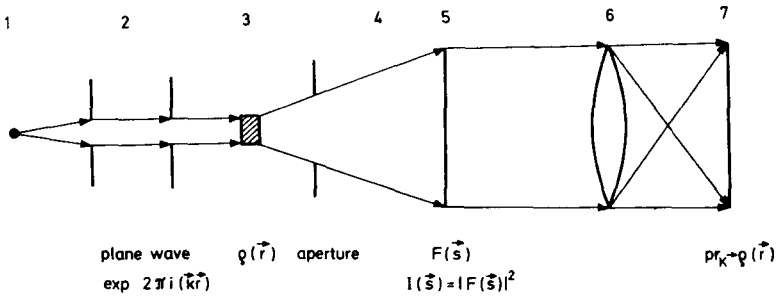

Fig. 1. Formation of the image in diffraction experiment
1 — radiation source, 2 — collimation and formation of the initial beam; **k** — the vector of the initial wave, 3 — object, 4 — diffraction, 5 — diffraction plane; vector $\mathbf{s}=\mathbf{k}-\mathbf{k}''$, $\mathbf{k}''$ is the vector of diffracted wave, 6 — optical system, 7 — image

performed, with some other degree of reliability and accuracy; to do this mathematically is the ultimate goal of this kind of research.

Let us consider this scheme in more detail going from left to right, i.e. from a radiation source to the image.

Various sources of radiation, X-rays, electrons and neutrons are used. All of them have wavelength values which are suitable for structural investigations: their wavelengths ranging from several tenths of nanometers down to several hundredths, which result in a resolution of atomic scales. However, due to a different nature of interaction with various substances, the "scattering matter" appears to be different for each of them: for X-rays the electron density, for electrons the electrostatic potential, and for neutrons the nuclear and spin density of the object. It should be noted that when we are not interested in the structure of an object at the level of atomic resolution, as in the case of SAS, this difference is not essential. However, it may become important combining diffraction data of one and the same object obtained from various radiations.

Passing to the scattering object itself, the character of its order determines, to a large extent, the possibility of deriving the structure from a diffraction experiment. With higher order the separation of the object image $F(\mathbf{s})$ in the reciprocal (Fourier) space

$$F(\mathbf{s})=\mathscr{F}[\rho(\mathbf{r})]$$

becomes more distinct in principle, the derivation of the structure of an "immobile" non-averaged object from diffraction data is more readily obtained. With decreasing the ordering, textures of different type are obtained whose ordering character may be described by some averaging operator B (Vainshtein [1]). In this case the intensities in the reciprocal space are mixed up being averaged according to the same operator and it is impossible to extract a pure module of the scattering amplitude $|F(\mathbf{s})|=\sqrt{I(\mathbf{s})}$ from

the given point in the reciprocal space. The intensities are the Fourier transforms of self-convolution of an object specifically averaged:

$$Q(\mathbf{r}) = \rho(\mathbf{r}) * \rho(-\mathbf{r}).$$

The general scheme is given in Fig. 2. In the case of SAS by biomolecules in solution a random orientation exists (B is a spherical averaging) and most difficulties in the structure determination are due to this reason.

With the question of the properties of the object the problem of the information inherent in a scattering pattern is closely connected. As it is clear from Fig. 2 one may pass over from the intensities to the structure $\rho(\mathbf{r})$ using either the amplitudes $F(\mathbf{r})$ or the distance function $Q(\mathbf{r})$. Table I illustrates the possibilities of deriving the structure of an object from experimental data. It demonstrates the principal role of X-ray structure analysis of crystals in the investigation of matter and, particularly, of biomolecules.

At present, this method yielding a resolution of 0.3—0.15 nm has been used for investigating more than 100 globular proteins having molecular weights of 10 000 to 300 000. This made it possible to establish the main regularities in the packing of polypeptide chains as well as to reveal the mechanism of some enzymatic reactions. The structure of transport RNA and the protein coat of two small spherical viruses have been also found (Blundell and Johnson [2]).

The structure of larger proteins as well as of high polymeric DNA and RNA at high resolution is still unknown. This may be associated with the very large dimensions of these molecules and with difficulties (or sometimes the impossibility) in obtaining such objects in the form of single crystals. The multicomponent systems such as ribosomes, membranes, viruses, chromatines and so on are to be regarded as the most complex biological objects in structural investigations.

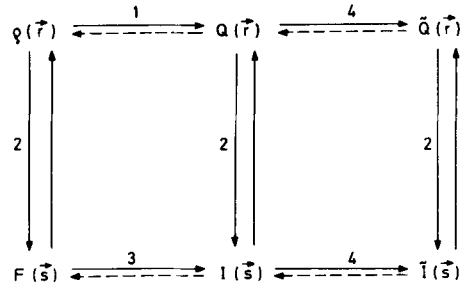

Fig. 2. Connections between real and reciprocal space patterns $\rho(\mathbf{r})$ — electron density, $F(\mathbf{s})$ — amplitude of scattering, $Q(\mathbf{r})$ — Patterson function, $I(\mathbf{s})$ — intensity of scattering, $\tilde{Q}(\mathbf{r})$ — Patterson function after averaging procedure, $\tilde{I}(\mathbf{s})$ — corresponding intensity of scattering, 1—4 — some mathematical treatments: 1 — self-convolution, 2 — Fourier-transform, 3 — multiplication by complex-conjugated value (square), 4 — averaging—transformation. Full arrows show the possibility of unambiguous transition. Dashed arrows show that the transition is not direct or ambiguous

Table I

The possibilities of the derivation of the structure from diffraction data

Methods and objects	Construction of $Q(\mathbf{r})$	Finding $F(\mathbf{s})$	Mathematical determination of phases	Physical measurement of phases	Physical image	Model building
X-ray diffraction						
Neutron diffraction						
Electron diffraction						
for crystals	+	+	$+^{(1)}$	$+^{(2)}$	–	$+^{(4)}$
for textures	+	(+)	$(+)^{(1)}$	–	–	$+^{(4)}$
for random systems	+	(+)	$(+)^{(5)}$	–	–	$+^{(4)}$
Electron-microscopy	–	–	–	+	$+^{(3)}$	–

+ possible, – impossible, (+) means the possibility in special cases or the possibility of ambiguous transition

1 — direct methods based on the relationships between modules of amplitudes; 2 — insertion of heavy atoms, dynamical scattering; 3 — holographic methods, calculation from an image; 4 — application of scattering invariants, various radiations, changes in contrast, minimizing functionals; 5 — direct method using spherical harmonics

Very valuable information on the structure of biological macromolecules has been obtained by means of electron microscopy (Vainshtein [3]). However, these studies are restricted by their own limitations. The space resolution presents the first difficulty. Despite the fact that the instrumental resolution of the modern electron microscopes is 0.15 nm biological samples must be prepared by staining with heavy-atom compounds; this reduces the resolution to 3—2 nm. The second limitation is the transition from an image which is a projection $p(\mathbf{r}_k) \rightarrow \rho(\mathbf{r})$ to the three-dimensional structure.

Thus the methods of X-ray diffraction and electron microscopy cannot be applied in all instances, especially in dealing with complex biological objects; with this type of samples one has to resort to SAS in solutions. Here the molecules are in their natural, native state, though due to spherical averaging a large amount of diffraction information is lost. Nevertheless many important characteristic features of an object may be found, and as it will be shown below, in some cases one can even directly determine the inner structure of a biomolecule.

The X-ray SAS owes its origin to the classical works of Guinier [4]. The name itself — SAS — is associated with the observation of diffuse scattering near the primary beam.

Although SAS has been applied for more than 40 years to many inhomogeneous systems such as coals, glasses, alloys, catalyzers and so on, the revival of the method is doubtlessly associated with biological objects.

The possibility of the study of dilute solutions of biomolecules is an impetus for the development of theoretical methods to extract structural information from scattering curves and for the development of new experimental methods. The success of X-ray small-angle studies of biological objects is associated with the development of X-ray experimental and calculation methods involving computers as well as the possibility of obtaining homogeneous preparations in large amounts.

In the last ten years the experimental SAS methods have, in principle, acquired new forms. New possibilities arose as a result of synchrotron radiation, powerful neutron beams, position-sensitive detectors. Let it be noted that the detailed investigation of the structure of biological objects, i.e. the search for the shape and inner structure of the molecules require precise intensity measurements at angles where the intensity of scattering by the solution decreases by 3—5 orders of magnitude as compared with the scattering at zero angle, and is only slightly different from the scattering by a pure solvent. Consequently, one of the main experimental requirements is the measurement of a wide range of small scattering angles. The position-sensitive detectors adopted from nuclear physics are rather suitable for this task.

Let us now consider the problem of deriving the structure from an experiment. The intensity of SAS (to be denoted by $I(s)$, where $s = 4\pi \sin \psi/\lambda$, λ = wavelength and 2ψ is the scattering angle) for any predetermined system may be calculated either analytically or by numerical methods. The problem of calculating the function $I(s)$ from the function $\rho(\mathbf{r})$ presents no difficulties. In principle, the classical Debye formula

(Guinier and Fournet [5])

$$I(s) = \int\int \rho(r)\,\rho(r')\frac{\sin s|\mathbf{r}-\mathbf{r}'|}{s|\mathbf{r}-\mathbf{r}'|}\,dv_1\,dv_2$$

is the basic equation and may be applied in most of the SAS cases.

However, when analysing an unknown object the investigator is naturally interested in the opposite problem, which consists in finding the distribution $\rho(\mathbf{r})$ from a given scattering intensity $I(s)$. This problem has no unambiguous solution due to considerable loss of information as a result of the spherical averaging. Up to now there have been no direct methods of structure determination based on SAS data, and therefore the principal method is still the comparison of experimental intensities (or their density distribution functions) with the scattering by model structures.

In the calculations of the scattering intensity many configurations from simple geometrical forms and their aggregates to compact and unfolded macromolecules, or particles with nonuniform scattering density distribution are considered. In some cases strict analytical solution can be obtained, however, in many cases only approximations with the aid of a computer are possible. The solution should fit the integral characteristics of an object, i.e. its invariants which are determined directly from the intensity: these are the radius of gyration R_g, the volume V, the surface S, the maximal size l_{max} and some others. These invariant quantities serve as the basis of a model both in the framework of the simplest geometrical bodies and in the homogeneous density approximation.

An important step forward was the development of the modelling method (Kratky [6]; Feigin [7]). We can now build up a model of a homogeneous body of a complicated form, filling its volume with certain elementary bodies, e.g. with cubes, balls, etc. By varying the form of the model one obtains different curves $I(s)$, which may be compared with the tentative form. It turned out that the theoretical SAS curves are rather sensitive to any changes in the model.

Thus the model method is a peculiar version of the trial-and-error method in classical structural crystallography. The choice of an initial model is, to some extent, an intuitive process, since any, even a very simple model of a molecule, is a function of many parameters. Here the superfluous details at the initial stage may only impede any subsequent research. The choice of an initial model (satisfying the invariants) is completed by outside information extracted from electron microscopy, biochemistry and physico-chemical data.

It is difficult to state unambiguously which of the numerous models is the most preferable one. The integral discrepancy factor may serve as an objective criterion. As a rule, one uses the following R factor

$$R_I = \frac{\int_{s_1}^{s_2} [I_{model}(s) - I(s)]^2 s^4\,ds}{\int_{s_1}^{s_2} I^2(s) s^4\,ds}, \qquad (1)$$

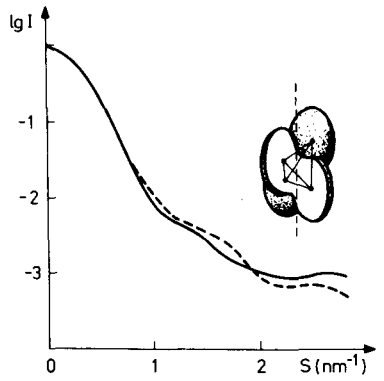

Fig. 3. The model of catalase
Full line — experimental curve, Dashed line — the curve of SAS by the model

where (s_1, s_2) is the interval whithin which the comparison is made. The derivation of the model from the SAS data may be compared with the search for a tentative model in the structure analysis of crystals.

As an example of applying the model method the SAS investigation of catalase is presented.

Catalase is an enzyme (mol. weight 250 000) which catalyzes the decomposition of hydrogen peroxide into water and oxygen. According to X-ray analysis and electron microscopy data catalase has the tetrahedral symmetry 222, which has been used as a starting structure to construct the model. Fig. 3 shows the best model found for catalase as well as its theoretical intensity curve together with the experimental curve. The model consists of four identical ellipsoids of revolution ($a = 1.7$ nm, $c = 4.7$ nm) spaced in the tetrahedral vertices with an edge length of 5.9 nm. The gap between the ellipsoids is also filled by scattering density (Vainshtein et al [8]).

More recent examples of the model method by the precise modelling of the form of a biomolecule may be represented by the studies of histidinedecarboxylase (Gonchar et al [9]) and the bacteriophage T7 (Rolbin et al [10]). The parametrisation of the model and the use of the discrepancy integrals will probably render it possible to find, in the future, an algorithm for looking for the "best" model by the method of non-local search or the least-square technique.

The objects with inhomogeneous density such as various nucleoproteins (viruses, ribosomes, etc.) are becoming important for small-angle investigations. With these samples the decisive role is played by the possibility of varying the solvent density which changes markedly the s dependence of I as well as the possibility of combining various radiations. The method known as "contrast variation" (Stuhrmann and Kirste [11]) allows after performing measurements in a series of solvents, the separation of three scattering functions

$$I(s) = I_F(s) + I_{FS}(s) + I_S(s),$$

where $I_F(s)$ is the scattering intensity corresponding to the homogeneous electron density, $I_S(s)$ is the scattering by inhomogeneities when the medium electron density is equal to the mean electron density of a particle, $I_{FS}(s)$ is the interference term, which may be both positive and negative. For biological objects the contrast variation is best achieved for neutron scattering with the aid of solvents of different concentration of heavy water.

Some useful information about the specific features of biomolecules can also be obtained with the aid of the recently developed methods of heavy-atom markers (Vainshtein et al [12]), anomalous SAS (Stuhrmann [13]), and the methods using real-space information (Glatter [14]).

It should be stressed, however, that all these methods, except a very few special cases, do not provide directly the structural information about a particle. Only some of its general parameters can be specified without modelling. The relationship between the theoretical and experimental possibilities in the small angle studies of biological objects is rather complicated: partly because up to now the experiment is devoid of sufficient accuracy, so that some parameters which may be found theoretically cannot be calculated from experimental data; and partly because even the high precision experimental scattering curves do not yield sufficient information; consequently, we still fail to provide direct information about the structure. So one of the most important problems in the field of SAS to solve biomolecular objects is the creation of a direct method to find their structure, i.e. to obtain the density distribution $\rho(\mathbf{r})$.

It is clear, however, that the task is to obtain a three-dimensional function $\rho(\mathbf{r})$ from the one-dimensional function $I(s)$, which has an infinite variety of solutions. This means that any direct method should be regarded as an approach which enables to narrow down this variety as much as possible by means of imposing some physically justified restrictions. We have developed the direct method of structure analysis in SAS using the information about the symmetry of a particle, its dimensions and the range of scattering density (Svergun et al [15]). The main features of the method as well as an example of its application to the concrete structure investigation are given below. To choose a convenient class of possible solutions the apparatus of spherical harmonics was used. Such mathematical approach was introduced in SAS by Stuhrmann [16]. If one represents the density $\rho(\mathbf{r})$ as a series

$$\rho(\mathbf{r}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \rho_{lm}(r) Y_{lm}(\theta, \varphi) = \sum_{l=0}^{\infty} \rho_l(\mathbf{r}), \tag{2}$$

where r, θ, φ are spherical coordinates, $Y_{lm}(\theta, \varphi)$ are spherical harmonics, $\rho_{lm}(r)$ are radial functions, and the equation

$$\rho_l(\mathbf{r}) = \sum_{m=-l}^{l} \rho_{lm}(r) Y_{lm}(\theta, \varphi) \tag{3}$$

defines the partial multipole densities, the intensity of the SAS is given by:

$$I(s)=2\pi^2\sum_{l=0}^{\infty}\sum_{m=-l}^{l}|A_{lm}(s)|^2\,. \tag{4}$$

The functions $\rho_{lm}(r)$ and $A_{lm}(s)$ are connected by the Hankel transform of order l:

$$A_{lm}(s)=i^l\sqrt{2/\pi}\int_0^{\infty}\rho_{lm}(r)j_l(sr)r^2\,\mathrm{d}r\,, \tag{5}$$

$$\rho_{lm}(r)=(-i)^l\sqrt{2/\pi}\int_0^{\infty}A_{lm}(s)j_l(sr)r^2\,\mathrm{d}r\,. \tag{6}$$

Here $j_l(sr)$ are spherical Bessel functions. Stuhrmann has shown that the function $I(s)$ is invariant to independent real space rotations of any partial density $\rho_l(\mathbf{r})$. So the manifold distributions $\rho(\mathbf{r})$ are determined by the same functions $\rho_{lm}(r)$ corresponding to the decomposition (4), and the set $\rho_{lm}(r)$ can be regarded as the variety of possible solutions introduced above.

It is clear that one can decompose $I(s)$ into a sum of squares of the amplitudes $A_{lm}(s)$, generally speaking, quite arbitrarily. Thus to use this approach it is nessesary to obtain the decomposition (4) corresponding to the real structure of a particle. Few attempts were made to solve the problem (Marguerie and Stuhrmann [17]; Stuhrmann and Fuess [18]), but the general algorithm has not been found yet.

Obviously, the summands cannot be separated from the sum unless some assumptions are made. First of all let us suppose that the density $\rho(\mathbf{r})$ is satisfactorily represented by a finite sufficiently small number of harmonics. Furthermore, as it results from formulae (5), functions $A_{lm}(s)$ are independent of the value of m upon the fixed functions $\rho_{lm}(r)$. This fact means that the contributions of harmonics with different m and the same l are in principle inseparable in the scattering intensity $I(s)$. Thus the formulae (4) can be rewritten:

$$I(s)=2\pi^2\sum_{l=0}^{\infty}|A_l(s)|^2\,, \tag{7}$$

where

$$|A_l(s)|^2=\sum_{m=-l}^{l}|A_{lm}(s)|^2\,. \tag{8}$$

In this way the problem consists of obtaining the decomposition (7), i.e. harmonics with different l values. Each term of the sum (7) determines according to transform (6) a radial function $\rho_l(r)$. In the case of an axially symmetrical particle ($\rho_{lm}(r)\equiv 0$ if $m\neq 0$) the sum over m vanishes and one has simply $\rho_l(r)=\rho_{l0}(r)$. In the general case the decomposition (7) holds, but the functions $\rho_l(r)$ have no explicit physical meaning, they are superpositions of the functions $\rho_{lm}(r)$.

Whatever the meaning of the functions $\rho_{lm}(r)$, there are strictly specified angular non-uniformities depending on the number of the harmonic l, inherent to any partial density (3), and therefore there are definite regions in the reciprocal space where the scattering intensity $|A_l(s)|^2$ contributes maximally to $I(s)$ (in accordance with the order of the Hankel transform (5)). Consequently, the changes of the relative contribution of different harmonics to the total intensity $I(s)$ are connected with their numbers.

Besides, the possible mode of $A_l(s)$ is restricted by the fact that any biomolecule is finite in space, i.e. there exists an R value so that

$$\rho(\mathbf{r}) \equiv \rho_{lm}(\mathbf{r}) \equiv 0, \text{ if } r > R. \tag{9}$$

The peculiarities of the behaviour of the functions $A_l(s)$ examined above can be regarded as the physical foundation for developing an algorithm to decompose the scattering intensity into functions (7).

Actually let us specify the $L+1$ functions $\rho_l^{(k)}(r)$, obeying the space restrictions and determining the functions $A_l^{(k)}(s)$ which in turn determine, according to (7), the intensity $I^{(k)}(s)$. This $I^{(k)}(s)$, generally does not coincide with the true intensity $I(s)$.

In order to fit the intensity $I(s)$, by keeping the relative contributions of different harmonics the same, one may renormalize the amplitudes as follows:

$$\tilde{A}_l^{(k)}(s) = A_l^{(k)}(s) \sqrt{\frac{I(s)}{I^{(k)}(s)}}. \tag{10}$$

The amplitudes $\tilde{A}_l^{(k)}(s)$ will obey Eq. (7), but the set $\tilde{\rho}_l^{(k)}(r)$ specified by $\tilde{A}_l^{(k)}(s)$ will generally not obey (9). Let us therefore set

$$\tilde{\rho}_l^{(k+1)}(r) = \tilde{\rho}_l^{(k)}(r) \Pi(r-R) = \begin{cases} \rho_l^{(k)}, & r \leqq R, \\ 0, & r > R. \end{cases} \tag{11}$$

The amplitudes of the next approximation $A_l^{(k+1)}(s)$ are determined by the function (11) according to (5).

Thus, with each step of the iterative process determined by Eqs (10) and (11) the partial amplitudes are redistributed in accordance with the number of the harmonic l, the value of R and the scattering intensity $I(s)$. Since no a priori information about the radial functions, except condition (9), is available, the step functions

$$\rho_l^{(0)}(r) = \Pi(r-R) = \begin{cases} 1, & r \leqq R \\ 0, & r > R \end{cases} \tag{12}$$

seem to be convenient for the first approximation.

It is possible to show that the procedures (10)–(11) converge to the true intensity $I(s)$. To estimate the deviations in reciprocal space between $I(s)$ and $I^{(k)}(s)$ the same R

factor has been used as for the model method. The convergence of the process over the R factor is uniform.

The efficiency of this procedure has been verified by means of model examples. The transforms (5)—(6) have been computed by the Simpson rule, the termination effects in (6) (the intensities have been calculated up to some finite value s_{max}) were reduced by the generalized Steklov filter (Rolbin et al [10]).

In the simplest case $L=0$ (spherically-symmetric particle) the problem is reduced to the determination of the true set of signs of the function $\sqrt{I(s)}$. When the change of signs stops, the process is interrupted. 5—6 iterations are usually quite enough to obtain the solution, function $\rho(r)$ is restored practically completely. This is an example of a quite unfavourable case of the multistep function $\rho(r)$ shown in Fig. 4. The restoring appears to be good, the R factor is comparable with one of the transforms by true signs, i.e. with the termination effects.

The most important practical case, however, corresponds to $L>0$. In this case the convergence of the process to the true intensity $I(s)$ does not guarantee the convergence to the true radial functions $\rho_l(r)$. The incorrectness of the problem (one should restore the summands from the sum of their squares) may lead to considerable distortions of these functions. To reduce them natural physical conditions were used,

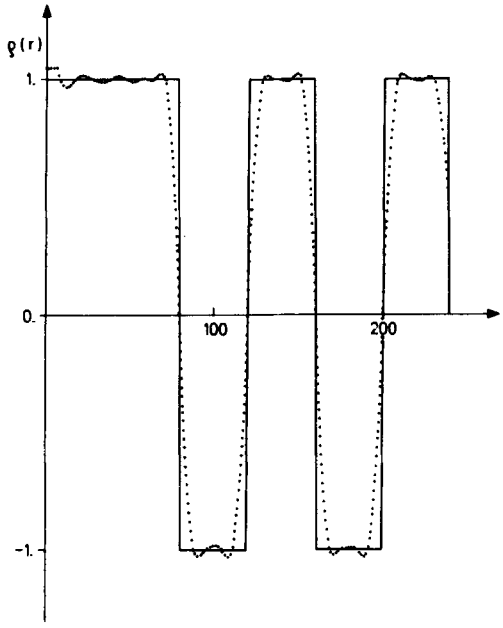

Fig. 4. The restoring of the structure of a spherically-symmetrical particle
Full line — true density distribution, Dotted line — restored distribution, Values $R=24$ nm, $s_{max}=5$ nm^{-1}; the result of the 7th iteration is shown, $R_f=2.5\times10^{-4}$

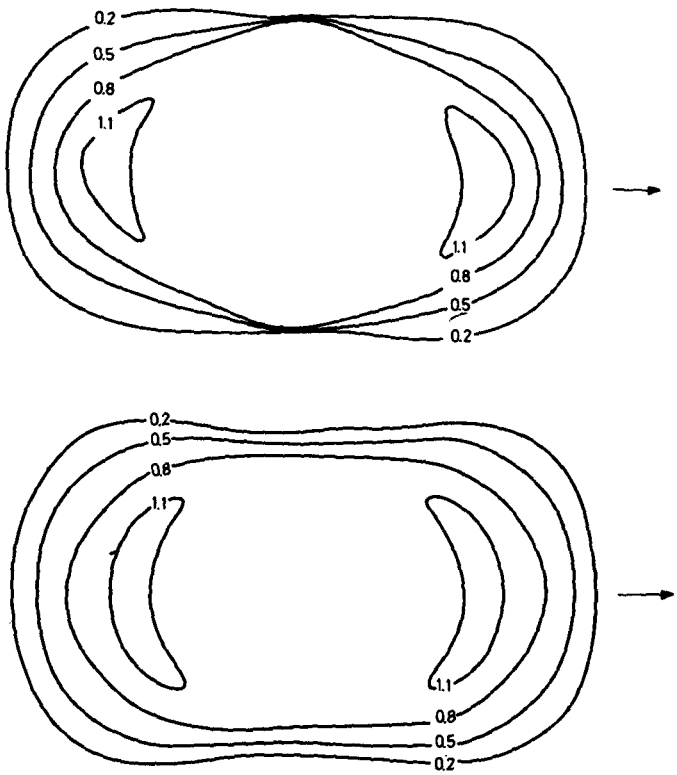

Fig. 5. The restoring of the model of an axially-symmetrical structure
a — true structure, b — restored structure
The distribution in the cross-section containing the axis of axial symmetry (indicated by an arrow) is shown; values $R=10\,\text{nm}$, $s_{\max}=5\,\text{nm}^{-1}$. The numbers refer to density levels in $10^{-3}\,\text{nm}^3$. The result of the 9th iteration is presented, $R_f=7.8\times10^{-3}$

namely the existence of limiting values of the scattering density

$$\rho_{\min}\leqq\rho(\mathbf{r})\leqq\rho_{\max}\,. \tag{13}$$

In fact, one can introduce the corrections $\rho_l(r)$ to the radial functions ensuring the correctness of (13) with a minimal R factor. It is convenient to express the corrections as a series of Laguerre polynomials (see Stuhrmann [16]). Assuming the corrections to be small enough, one can solve the problem by means of linear programming (Dantzig [20]). We have used the modified simplex method (ibid). It should be noted that the decreasing of the R factor suggests the relevance of this procedure.

An example of using procedures (10)—(11) with the restrictions (13) is presented in Fig. 5. The axially-symmetrical model structure (Fig. 5a) is constructed from three first non-zero radial functions of an ellipsoid of revolution with the density $\rho(\mathbf{r})=1$,

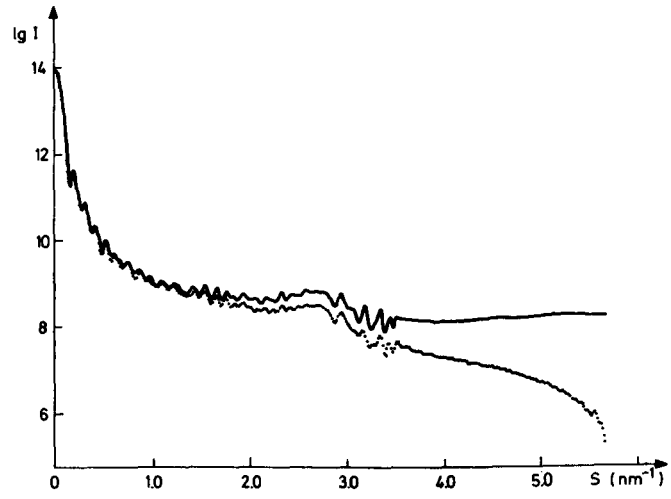

Fig. 6. The scattering curves of bacteriophage T7
Full line — experimental curve, Dotted line — the SAS curve from the restored structure

half-axes $a = 5$ nm, $c = 10$ nm. The structure shown in Fig. 5b is constructed from the intensity calculated for the previous one. The agreement seems to be quite satisfactory.

The examples shown above illustrate the efficiency of the procedure using parameters R, ρ_{min}, ρ_{max} and a set of exactly defined harmonics. In practice, these parameters are as a rule known with some errors. We have verified the stability of the procedure with respect to the deviation of these input parameters in the model examples. It has been proved that the method yields suitable solutions with deviations up to 10% in R and 20% in ρ_{min} and ρ_{max}. The errors in the set of harmonics may be connected mainly with the fact that any particle is generally represented by an infinite series of harmonics, whereas one has only a finite number of terms. It may be noted, however, that the spherical harmonics constitute a complete system of functions, and a few first terms in the series would describe the whole structure, as a rule, quite sufficiently.

It should be stressed that the procedure holds also for particles with arbitrary density distribution. In the general case in order to select a class of possible solutions one should perform a new separation, the separation of functions $\rho_{lm}(r)$ from $\rho_l(r)$, with the aid of some additional (non SAS) informations. Selecting one (or several) relevant solutions one should analyze independent rotations of different partial structures $\rho_l(\mathbf{r})$ (the problem is much easier in the axially-symmetrical case since any rotation of the partial density other than π violates the given axial symmetry).

As an example of a practical application of the method to certain structural investigation some results are presented which were obtained from the SAS curve of the

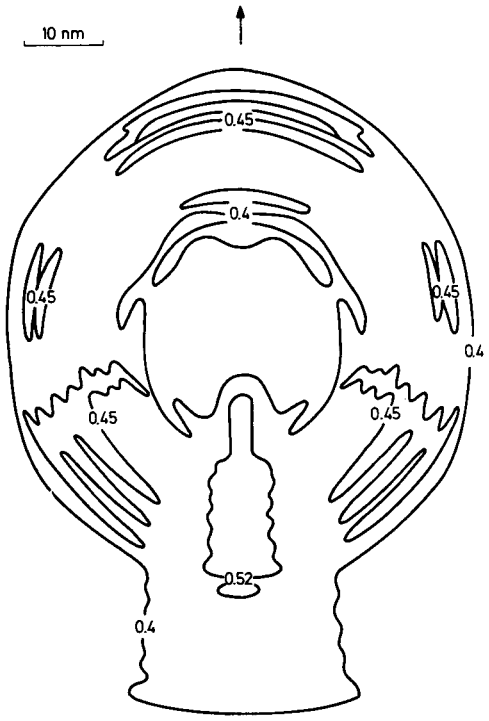

Fig. 7. The map of electron density of bacteriophage T7. The cross-section containing the axis of axial symmetry (indicated by an arrow) is shown. Level 0.4 corresponds to the protein, 0.45 — to strongly hydrated DNA, 0.52 — slightly hydrated DNA. The numbers refer to density levels in 10^{-3} nm^3. The result of the 8th iteration is given, $R_I = 2.2 \times 10^{-2}$

bacteriophage T7. This large bacterial virus has been studied also by various physical and chemical methods. It possesses an approximately axially-symmetrical constitution (isometric polyhedral head and cylindrical tail). Highly refined solutions of the phage T7 for our SAS studies were extracted at the Semmelweis Medical University, Budapest, under the guidance of Prof. I. Tarján and Prof. G. Rontó. We recently measured the X-ray SAS curve of the phage quite precisely in a wide range of angles, a number of general parameters of the phage have been determined and its model has been suggested (Rolbin et al [19]). We have applied the direct method to interpret the part of the curve shown in Fig. 6. The parameters R, $\rho_{\min}$ and $\rho_{\max}$ have been selected by using general phage parameters determined from this curve. An iterative procedure has been applied and the result is presented in Fig. 7.

This figure is a map of the electron density of the phage T7 in cross-section containing the axes of axial symmetry, the radial resolution is 1.2 nm. One can see the projection of the phage head to be embraced by a six-fred symmetry with the corresponding edges equal to 35 nm. The tail appears as a circular cylinder with a radius 11 nm and a height of 18 nm. This is a protein core in the central part of the head

with a diameter of about 24 nm, in good agreement with the data of neutron scattering (Agamalian et al [21]). There is a cylindrical region of higher DNA concentration near the tail, whereas the phage DNA as a whole is strongly hydrated leading to an almost uniform density within the phage. Further on, there are hints in favour of a regular arrangement of DNA, which is revealed by the arcs in Fig. 7.

The scattering curve of the restored density distribution (Fig. 6) coincides with the experimental one well enough up to the scattering angles responsible for the dimensions in a particle compared with the value of radial resolution. All T7 features specified are in good accord with the data obtained (as a rule, by indirect methods) from other physical and chemical techniques.

The results suggest that this method will be quite useful for the structure analysis of biological macromolecules in solution by SAS studies.

It is a honour for the authors to publish this paper in the issue devoted to the 70th birthday of Prof. I. Tarján, who made valuable contributions both to solid state physics as well as biophysics. The results presented in this paper on bacterial viruses would have been impossible without our collaboration with Prof. I. Tarján and his coworkers.

References

1. B. K. Vainshtein, Diffraction of X-rays by Chain Molecules, Elsevier Publ. Comp., Amsterdam, 1966.
2. T. L. Blundell and L. N. Johnson, Protein Crystallography, Acad. Press, New York, 1976.
3. B. K. Vainshtein, Advances in Optical and Electron Microscopy, **7**, 281, 1978.
4. A. Guinier, Compt. Rend., **206**, 1374, 1938; **206**, 1641, 1938.
5. A. Guinier and G. Fournet, Small-angle Scattering of X-rays, J. Wiley, New York, 1955.
6. O. Kratky, Progress in Biophysics, **13**, 105, 1963.
7. L. A. Feigin, Kristallografia, **16**, 711, 1971.
8. B. K. Vainshtein, S. Ya. Karpuhina, N. I. Sosfenov and L. A. Feigin, Doklady AN SSSR, **207**, 1336, 1972.
9. N. A. Gonchar, Yu. M. Lvov, T. G. Samsonidze, L. A. Semina and L. A. Feigin, Biofizika, **23**, 768, 1978.
10. Yu. A. Rolbin, D. I. Svergun, L. A. Feigin and B. M. Schedrin, Kristallografia, **25**, 1125, 1980.
11. H. B. Stuhrmann and R. G. Kirste, Z. Phys. Chem., **46**, 247, 1965.
12. B. K. Vainshtein, L. A. Feigin, Yu. M. Lvov, R. I. Gvozdev, S. A. Marakushev and G. I. Likhtenshtein, FEBS Letters, **116**, 107, 1980.
13. H. B. Stuhrmann, Kristallografia, **26**, 956, 1981.
14. O. Glatter, J. Appl. Cryst., **12**, 166, 1979; **14**, 101, 1981.
15. D. I. Svergun, L. A. Feigin and B. M. Schedrin, Doklady AN SSSR, **261**, 887, 1981; Kristallografia, **26**, 1171, 1981.
16. H. B. Stuhrmann, Acta Crystallogr., **A26**, 297, 1970.
17. G. Marguerie and H. B. Stuhrmann, J. Mol. Biol., **102**, 143, 1976.
18. H. B. Stuhrmann and H. Fuess, Acta Crystallogr., **A32**, 67, 1976.
19. Yu. A. Rolbin, D. I. Svergun, L. A. Feigin, Sh. Gashpar and G. Ronto, Doklady AN SSSR, **255**, 1497, 1980.
20. G. B. Dantzig, Linear Programming and Extensions, Princeton Univ. Press, New Jersey, 1963.
21. M. M. Agamalian, G. M. Drabkin, A. A. Dovgikov, T. I. Krivshich, Yu. M. Lvov, L. A. Feigin, G. Ronto and Sh. Gashpar, Kristallografia, **27**, 92, 1982.