Acta Cryst. (1991). A47, 736-744

New Developments in Direct Shape Determination from Small-Angle Scattering. 1. Theory and Model Calculations

BY D. I. SVERGUN* AND H. B. STUHRMANN

GKSS Research Centre, GKSS-WS, 2054 Geesthacht, Germany

(*Received* 28 *March* 1991; *accepted* 30 *May* 1991)

Abstract

Improvements in the shape determination method based on multipole expansion [Stuhrmann (1970), *Z. Phys. Chem.* (*Frankfurt am Main*), **72**, 177-198] are described. Fast algorithms for evaluating shape scattering from the multiple coefficients are developed. The relationship between the resolution in real and reciprocal space is established. The improved technique is verified by means of model calculations. It is shown that the low-resolution shape estimation can be performed as a straightforward procedure. Shape refinement at higher resolution is ambiguous; a variety of shapes can be generated which fit the given intensity curve neatly. A criterion is given to select the most plausible solution in higher-resolution studies.

1. Introduction

The possibility to extract structural information from small-angle scattering (SAS) data depends to a large extent on the nature of the system under investigation. SAS methods can be applied to the study of a wide range of non-crystalline objects and some general parameters characterizing the sample can be evaluated directly from the experimental data (Feigin & Svergun, 1987). One of the most favorable cases is that of monodisperse systems consisting of randomly oriented particles, of which dilute solutions of biological macromolecules are good practical examples. In this case the SAS intensity is proportional to the scattering from a particle averaged over all orientations and this allows one readily to obtain some geometrical and physico-chemical parameters of the particle from the scattering curve.

As a result of the spherical averaging of the scattering data a considerable amount of information is lost and the relationship between the particle structure and the SAS intensity is not unequivocal. Restrictions must be imposed on possible solutions in order to extract information about the three-dimensional particle structure from the one-dimensional SAS curve. The assumption that one is dealing with homogeneous particles for which only the shape is to be determined is especially important in practice. The SAS intensity corresponding to the shape scattering can be obtained experimentally using contrast variation methods (Stuhrmann & Kirste, 1965).

Simple approximations (*e.g.* three-axial ellipsoids, prisms *etc.*) have been used for many years to characterize the particle shape (see, for example, Glatter & Kratky, 1982). The problem of more detailed shape description is normally handled in terms of models. Different approaches have been proposed to evaluate the SAS curves from models (Rolbin, Kayushina, Feigin & Schedrin, 1973; Müller, Damaschun & Hübner, 1979; Glatter, 1980; Müller, 1983). In this paper we deal with direct shape determination procedures based on the multipole expansion.

2. Shape determination using multipole expansion

First we shall briefly outline the formalism of the shape determination using multipole expansion. The approach is described in detail in the original papers (Stuhrmann, 1970*a*, *b*).

A three-dimensional function $\rho(\mathbf{r})$ (particle density distribution) can be represented as a series

$$\rho(\mathbf{r}) \simeq \rho_L(\mathbf{r}) = \sum_{l=0}^{L} \sum_{m=-l}^{l} \rho_{lm}(r)\, Y_{lm}(\omega), \qquad (1)$$

where $(r, \omega) = (r, \theta, \varphi)$ are spherical coordinates,

$$\rho_{lm}(r) = \int \rho(\mathbf{r})\, Y^*_{lm}(\omega)\, \mathrm{d}\omega \qquad (2)$$

are the radial functions and $Y_{lm}(\omega)$ are spherical harmonics. Here the truncation value L describes the resolution, *i.e.* accuracy of the representation of the particle structure [$\rho_L(\mathbf{r}) \to \rho(\mathbf{r})$ when $L \to \infty$]. With this expansion the particle SAS intensity is expressed as (Harrison, 1969; Stuhrmann, 1970*a*)

$$I(s) = \sum_{l=0}^{L} \sum_{m=-1}^{l} |A_{lm}(s)|^2, \qquad (3)$$

where s is the modulus of the scattering vector $\mathbf{s}$ [$s = (4\pi/\lambda) \sin\theta$, λ is the wavelength and 2θ is the scattering angle] and

$$A_{lm}(s) = i^l (2/\pi)^{1/2} \int_0^\infty \rho_{lm}(r) j_l(sr) r^2\, \mathrm{d}r \qquad (4)$$

* On leave from Institute of Crystallography, Academy of Sciences of the USSR, Leninsky prospekt 59, 117333 Moscow, USSR.

0108-7673/91/060736-09$03.00

are the Hankel transforms of the radial functions where $j_l(sr)$ are spherical Bessel functions.

The structure of a wide variety of homogeneous particles can be described with the help of the angular shape function $F(\omega)$ as

$$\rho(\mathbf{r}) = \begin{cases} 1 & 0 \le r < F(\omega) \\ 0 & r \ge F(\omega). \end{cases} \tag{5}$$

This function can also be developed into the series

$$F(\omega) = \sum_{l=0}^{L} \sum_{m=-l}^{l} f_{lm} Y_{lm}(\omega) \tag{6}$$

where the multipole coefficients are complex numbers

$$f_{lm} = \int F(\omega)\, Y^*_{lm}(\omega)\, \mathrm{d}\omega. \tag{7}$$

The set of f_{lm} coefficients describes the shape of the particle at the given resolution. Representing the spherical Bessel function as a power series

$$j_l(sr) = \sum_{p=0}^{\infty} [(-1)^p / 2^p p! [2(l+p)+1]!!](sr)^{l+2p} = \sum_{p=0}^{\infty} d_{lp}(sr)^{l+2p} \tag{8}$$

and substituting this series into (4), one obtains (Stuhrmann, 1970*b*)

$$A_{lm}(s) = i^l (2/\pi)^{1/2} \times \sum_{p=0}^{\infty} [d_{lp} f^{(l+2p+3)}_{lm} / (l+2p+3)] s^{l+2p}, \tag{9}$$

where

$$f^{(q)}_{lm} = \int [F(\omega)]^{(q)}\, Y^*_{lm}(\omega)\, \mathrm{d}\omega. \tag{10}$$

Using (3), the SAS intensity is represented as a power series

$$I(s) = \sum_{n=0}^{\infty} s^{2n} a_n = \sum_{n=0}^{\infty} s^{2n} \sum_{l=0}^{L} \sum_{p=0}^{n-l} \sum_{m=-l}^{l} d_{lp} f^{(k_1)}_{lm} d_{l,n-l-p} f^{(k_2)}_{lm} / k_1 k_2, \tag{11}$$

$k_1 = l+2p+3$, $k_2 = 2n-l-2p+3$. Therefore, the coefficients of the power series of the SAS intensity (which can be extracted from the experimental data) are expressed as a nonlinear combination of the multipole coefficients $f^{(q)}_{lm}$ of the shape function. These relationships can be used to find the shape by minimizing the deviations between the observed and calculated coefficients a_n.

This method was shown to be useful for low-resolution shape determination when using only the few first terms in series (11) and, correspondingly, a restricted number of multipoles in expansion (6). In particular, a model of the 50S ribosomal subunit at resolution up to $L=3$ was proposed (Stuhrmann, Koch, Parfait, Haas, Ibel & Crichton, 1977).

3. New features

The formalism above did not find wide application, mainly due to difficulties in its practical implementation and, in particular, the severe problem of evaluation of the coefficients $f^{(q)}_{lm}$ for higher l and q values. The higher l and q, the less reliable the numerical integration in (10). Series (11) becomes divergent with increasing s and its truncation restricts the possibilities of representing the SAS intensity. In practice one can only use about a dozen coefficients and terms in the power series. This means that in this form the method is restricted to a resolution not better than $L=3$. Below we present improvements in the procedure which significantly increase the reliability of the technique and allow one to move to higher-resolution studies.

Let us substitute in (10) $F^{(q)}(\omega) = F(\omega) F^{(q-1)}(\omega)$ and represent $F(\omega)$ according to (6):

$$f^{(q)}_{lm} = \int [F(\omega)]^{(q-1)} \sum_{k=0}^{L} \sum_{t=-k}^{k} f_{kt} Y_{kt}(\omega)\, Y^*_{lm}(\omega)\, \mathrm{d}\omega. \tag{12}$$

Using the following formula for the product of two spherical harmonics (Edmonds, 1957)

$$Y_{lm}(\Omega) Y_{kt}(\Omega) = \sum_{p=|l-k|}^{l+k} [(2l+1)(2p+1)(2k+1)/4\pi]^{1/2} \times \begin{pmatrix} l & k & p \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l & k & p \\ m & t & -m-t \end{pmatrix} Y^*_{p,-m-t}(\omega) \tag{13}$$

where

$$\begin{pmatrix} l & k & p \\ m & t & -m-t \end{pmatrix}$$

are $3j$ Wigner symbols, integrating over ω and taking into account the orthogonal properties of spherical harmonics, one obtains the following recurrence formula:

$$f^{(q)}_{lm} = (-1)^m \sum_{k=0}^{L} \sum_{p=|k-l|}^{k+l} [(2l+1)(2k+1) \times (2p+1)/4\pi]^{1/2} \begin{pmatrix} l & k & p \\ 0 & 0 & 0 \end{pmatrix} \times \sum_{t=-k}^{k} f_{kt} f^{(q-1)}_{p,m-t} \begin{pmatrix} l & k & p \\ m & t & m-t \end{pmatrix}. \tag{14}$$

This formula allows one to evaluate the coefficients $f^{(q)}_{lm}$ for $q>1$ without integration.

The representation of the SAS intensity is much simplified by first evaluating the amplitudes $A_{lm}(s)$. The convergence of series (9) is obviously twice as fast as that of series (11), that is, twice as large an angular region can be taken into account without

producing numerical instability. The amplitudes having been evaluated, the intensity can be summed according to (3). The functional to be minimized then has the form

$$\Phi = \Phi[\{f_{lm}\}] = \int_{s_{min}}^{s_{max}} |I_{exp}(s) - I_{mod}(s)| W(s)\, ds \times \left(\int_{s_{min}}^{s_{max}} I_{exp}(s) W(s)\, ds \right)^{-1} \quad (15)$$

where the fitting region and the choice of the weighting function $W(s)$ depend on the resolution which one can obtain, the accuracy of the experiment *etc.*

With these improvements both higher harmonics and a wide angular range of the SAS intensity curve can be taken into account, in principle allowing higher-resolution studies.

4. Resolution in real and reciprocal space

To show the possibilities of the improved technique, a series of model calculations has been performed. We plan to apply the shape determination in the study of the ribosome structure, therefore the two electron microscope models of the 50S ribosomal subparticle are used as model objects. One of them has been proposed by Lake (1976), the other by Yonath, Leonard & Wittmann (1987). Shape coefficients have been calculated using the solid model in the first case and the cross sections of the model in the second case (Yonath, 1989). The models constructed from the f_{lm} sets with a resolution up to $L=7$ are presented in Figs. 1(*a*), (*b*); the corresponding SAS curves are shown in Fig. 1(*c*) (they are presented in the angular region where the shape scattering curve can be reliably determined in practice).

First, it is necessary to establish the relationship between the resolutions in real and reciprocal space. The former is determined by the maximum number of harmonics L, the latter by the maximum momentum transfer s_{max}. A simple estimation of the resolutions can be made as follows. Let R_0 be an average particle radius (say the radius of the sphere with the same volume). Then the resolution provided by the lth harmonic is $2\pi R_0/(l+1)$. On the other hand, the Bragg resolution is $2\pi/s$. Comparing the two expressions, one obtains that, for a multipole resolution L, the portion of the scattering curve up to at least $s_{max} = (L+1)/R_0$ must be available.

A similar estimation follows from the sampling theorem (Shannon & Weaver, 1949; Taupin &Luzzati, 1982) if one considers each new term in the sum over l in series (6) as an independent parameter: $s_{max} = \pi(L+1)/D_{max}$, D_{max} being the maximum particle diameter. It is worth noting here that the number of f_{lm} coefficients describing the particle shape is, of course, much higher than $L+1$, but they cannot be considered as 'independent parameters'. For example, it is easily shown that

$$\int [F(\omega)]^2\, d\omega = \sum_{l=0}^{L} F_l^2 = \sum_{l=0}^{L} \sum_{m=-l}^{l} |f_{lm}|^2 \quad (16)$$

and, therefore, each partial sum F_l^2 is conserved under arbitrary rotations of the particle.

The estimations are illustrated in Figs. 2(*a*), (*b*), where the relative multiple contributions

$$i_l(s) = I_l(s)/I(s); \quad I_l(s) = \sum_{m=-l}^{l} |A_{lm}(s)|^2 \quad (17)$$

are shown for the two models (the R_0 values are 86·4 and 84·9 Å for Lake's and Yonath's models, respectively). One can see that the angular range up to $s \simeq 0{\cdot}06$ Å^{-1} contains the main information about the harmonics with $l \leq 3$; to improve the resolution up to $L=5$, the region up to $s \simeq 0{\cdot}09$ Å^{-1} has to be taken into account. On the other hand, higher harmonics play no significant role at lower momentum transfers: for example, the partial intensity $I_7(s)$ is negligibly small for $s < 0{\cdot}06$ Å^{-1}.

The parameter R_0 is also important as a scale factor for numerical procedures. The transformation $r \rightarrow$

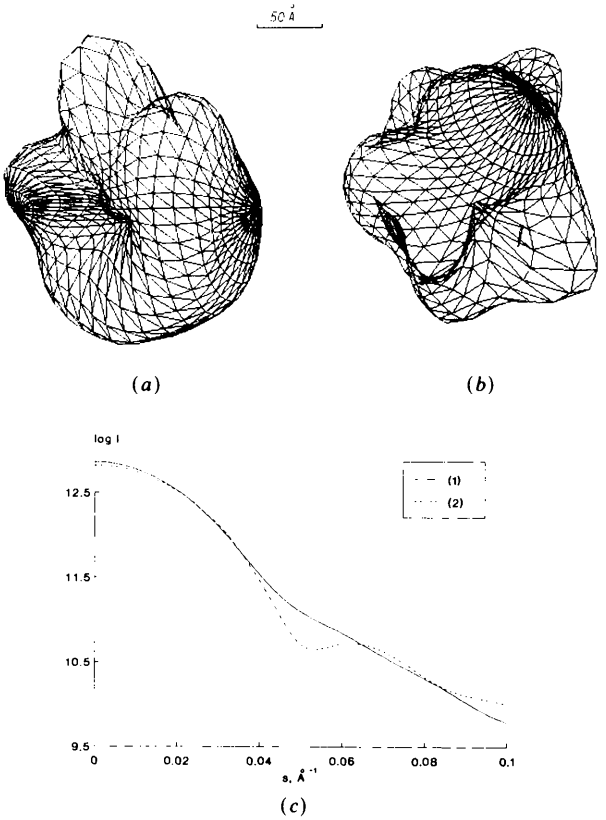


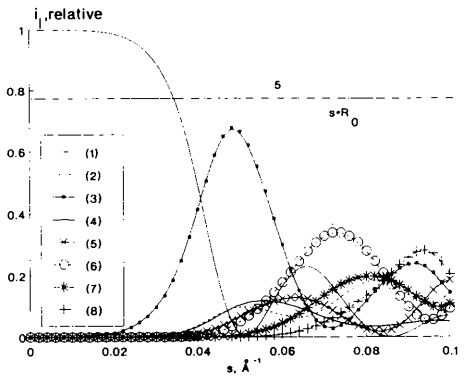
Fig. 1. The models of (*a*) Lake and (*b*) Yonath up to the resolution of $L=7$ and (*c*) their SAS curves: curve (1) corresponds to (*a*), (2) to (*b*).

r/R_0, $s \to sR_0$ has to be made in order to avoid rounding errors while evaluating $f_{lm}^{(q)}$ and $A_{lm}(s)$ according to (14) and (9). With such scaling, scattering curves up to $sR_0 = 13$-14 at the resolution $L = 7$-8 can be evaluated, which is sufficient for most practical purposes. Note that with the representation (10)-(11) only the range $sR_0 < 7$-8 at the resolution $L = 3$-4 could be covered.

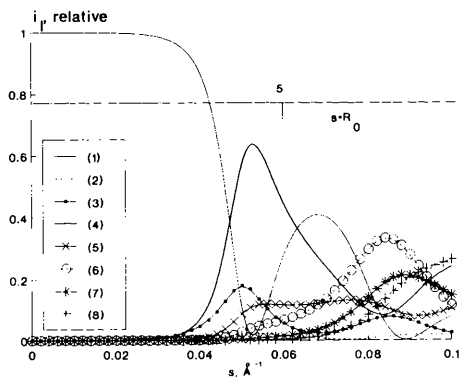
5. Model calculations at low resolution

The algorithm presented above was implemented in a Fortran program package. For minimizing functional (15), the optimization program *OPTIS*, developed at the Institute of Crystallography, USSR Academy of Sciences, was used. All the model calculations were performed on IBM-PC and VAX computers.

First, the method was tested under ideal conditions. That is, the model scattering curves were evaluated according to (9) and (3) using the series up to $L = 3$ and the shape functions were restored using the same resolution.



(a)



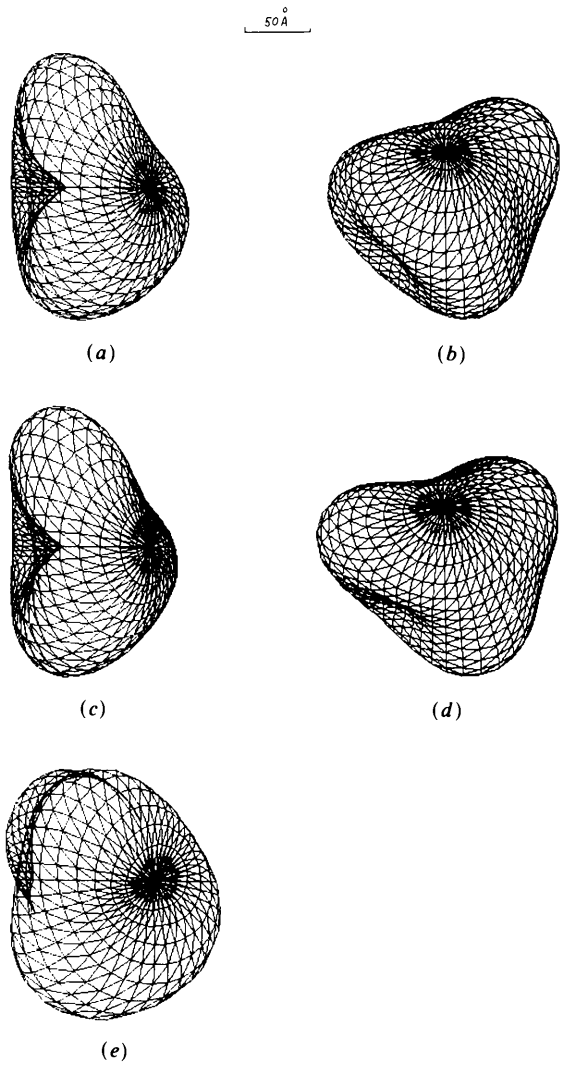
(b)

Fig. 2. Relative multipole contributions to the total scattering intensity for (*a*) Lake's and (*b*) Yonath's models. Curves (1)-(8) correspond to $l = 0$-7.

The models up to a resolution of $L = 3$ are shown in Figs. 3(*a*), (*b*); the corresponding scattering curves are given in Figs. 4(*a*), (*b*). One can see that the shapes as well as the scattering curves already differ significantly at this resolution. Our aim was to show that the method is able to distinguish between the two models and to restore the shape functions without any additional information.

As usual for minimization techniques, the problems of initial approximation, the type of the functional and the method of minimization arise. Various calculations have been made and the most effective shape restoration procedure can be formulated as follows:

(i) A simple but reasonable model can be taken as the initial approximation. We found that it is sufficient to use a three-axial ellipsoid approximation, where



(*a*) (*b*) (*c*) (*d*) (*e*)

Fig. 3. The models of (*a*) Lake and (*b*) Yonath up to the resolution of $L = 3$ and their restored shapes, (*c*) and (*d*), respectively. The shape (*e*) is obtained when restoring Lake's model starting from a sphere.

the half-axes are calculated so as to fit the three invariants of the SAS intensity curve (radius of gyration, volume, maximum diameter).

(ii) The most natural way of comparing the scattering curves *via* these Porod plots is also the best. That is, the weighting function $W(s) = s^2$ is to be taken in functional (15).

(iii) The combination of variable metrics and downhill simplex methods of optimization (Gill, Murray & Wright, 1981) proved to be a reliable and sufficiently fast minimization procedure.

The results of the model calculations (restored structures and restored scattering curves) are shown in Figs. 3 and 4. In both cases the starting point was a three-axial ellipsoid with the axial ratio $a:b:c = 0{\cdot}75:1:1{\cdot}4$; the angular range $0 < s < 0{\cdot}06$ Å^{-1} was used in functional (15).

It is worth noting here that the coefficients $f_{lm}^{(q)}$ are directly related to some important structural parameters of the particle, namely:

particle volume: $V = (4\pi)^{1/2} f_{00}^{(3)}/3$;

radius of gyration: $R_g^2 = 3f_{00}^{(5)}/5f_{00}^{(3)}$;

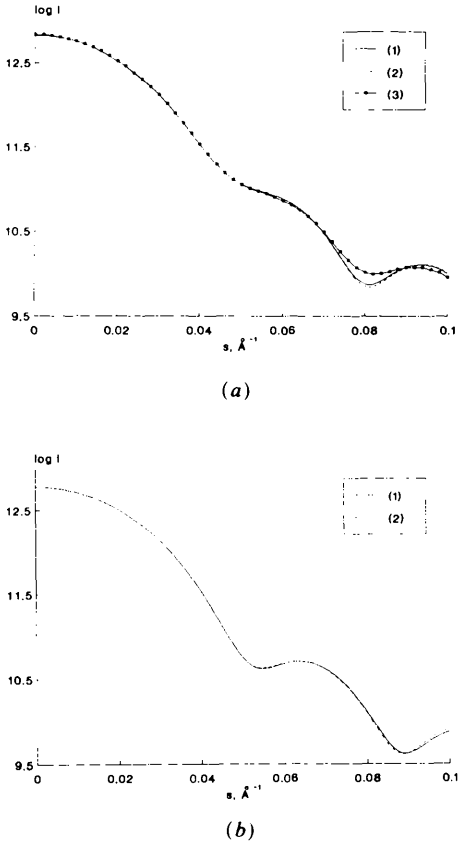


Fig. 4. (*a*) Scattering curve from Lake's model up to $L = 3$ (1) and its approximations by the models shown in Fig. 3(*c*) (2) and Fig. 3(*e*) (3); (*b*) scattering curve from Yonath's model up to $L = 3$ (1) and its approximation by the model shown in Fig. 3(*d*) (2).

coordinates of the center of mass:

$$\mathbf{r}_0 = (-6^{1/2}\,\mathrm{Re}\{f_{11}^{(4)}\}/4f_{00}^{(3)},\ 6^{1/2}\,\mathrm{Im}\{f_{11}^{(4)}\}/4f_{00}^{(3)},\ 3^{1/2}f_{10}^{(4)}/4f_{00}^{(3)}).$$

It is possible to control these parameters during the minimization procedure. In particular, to keep the particle reasonably close to the origin, the term $\mu|\mathbf{r}_0|$ ($\mu \simeq 10^{-4}R_0$) has been added to the functional (15). No restrictions were, however, imposed on the particle orientation, although this could be done by fixing some f_{lm}'s. For comparison with the initial shapes, the resulting structures were appropriately rotated. The rotation of the structure described by a set of f_{lm} coefficients by the Euler angles α, β, γ leads to the new set g_{lm} which can be easily calculated using the relation

$$g_{lm} = \sum_{k=-l}^{l} \mathcal{D}_{km}^{(l)}(\alpha, \beta, \gamma) f_{lk} \qquad (18)$$

where $\mathcal{D}_{km}^{(l)}(\alpha, \beta, \gamma)$ are matrix elements of the rotational operator (Edmonds, 1957).

The results are shown in Figs. 3 and 4. The minima of the functional are 2×10^{-4} for Lake's model and 9×10^{-4} for Yonath's model. The scattering curves are well restored even beyond the angular region used. The restoration of the models themselves is also satisfactory. The figure of merit for comparison, according to (16), is

$$M_L = \left[\frac{\sum_{l=0}^{L} \sum_{m=-l}^{l} [f_{lm}^{\mathrm{exp}} - f_{lm}^{\mathrm{mod}}]^2}{\sum_{l=0}^{L} \sum_{m=-l}^{l} [f_{lm}^{\mathrm{exp}}]^2} \right]^{1/2} = \left[\frac{\int [F^{\mathrm{exp}}(\omega) - F^{\mathrm{mod}}(\omega)]^2 \,\mathrm{d}\omega}{\int [F^{\mathrm{exp}}(\omega)]^2 \,\mathrm{d}\omega} \right]^{1/2}. \qquad (19)$$

Here exp refers to the model used for simulations, mod to the restored structure. In the present case, $M_3 = 0{\cdot}067$ for Lake's model and 0·073 for Yonath's model.

Similar results can be obtained starting with other reasonable approximations. Fig. 3 illustrates an average result rather than the best restoration. In particular, the shapes are better restored if one increases $s_{\max}$ (the restoration is almost perfect for $s_{\max} = 0{\cdot}1$ Å^{-1}). However, we tried to keep the range as small as possible taking into account further problems with resolution.

Starting with a less-reasonable initial shape one always runs the risk of being trapped in a local minimum of the functional which is remote from the true solution. This is illustrated in Fig. 3(*e*). This shape, which is a poor approximation of Lake's model ($M_3 = 0{\cdot}193$), is obtained when starting from a sphere. The value of the functional is much higher (7×10^{-3}), but the fit in the region $s < 0{\cdot}06$ Å^{-1} is still acceptable

[see curve (3) in Fig. 4(*a*)], especially taking into account possible errors in the experimental data. Indications that this solution is incorrect are found in the outer part of the scattering curve.

6. Termination effects and refinement

The previous results prove that the method can restore simple (but non-trivial) shape functions using a limited range of momentum transfer under 'ideal' conditions. That is, limited resolution of the multipole expansion did not introduce any errors. In reality, the scattering curves contain an infinite number of harmonics and the resolution effects have to be taken into account for the method to be of practical use.

The difficulties arising from the resolution effects are illustrated in Fig. 5, where the scattering curve of Lake's model without truncation [curve (1), the same as curve (1) in Fig. 1(*c*)] is shown together with the one truncated at $L=3$ [curve (2), the same as curve (1) in Fig. 4(*a*)]. The curves differ significantly and application of the same procedure as described above results in curve (3), which is in poor agreement with both 'truncated' and total intensities. The restored structure (figure of merit $M_3=0{\cdot}139$) deteriorates by the artificial increase of the multipoles with $l=2$ and especially $l=3$, resulting from the contributions of all higher multipoles (termination effects).

A trivial way to overcome this problem would be to take into account as many multipoles as possible in order to reduce the termination effects. For globular structures, the convergence of series (3) is fast and the resolution up to, say, $L=5$-6 is quite sufficient for experimentally measurable ranges of momentum transfer. This approach is, however, not acceptable from the numerical point of view. The number of f_{lm} coefficients to be adjusted is already high for $L=3$ (17 coefficients); it increases to 36 for $L=5$ and to 64 for $L=7$. Moreover, the higher harmonics overlap

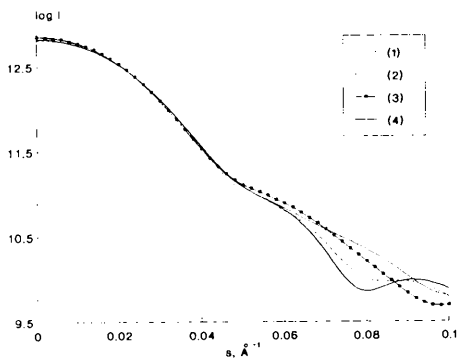


Fig. 5. Scattering curve from Lake's model without truncation of the multipole expansion (1); truncated at $L=3$ (2); obtained from the shape restored using the truncated at $L=3$ multipole expansion and curve (1), $s_{max}=0{\cdot}06$ Å^{-1} (3); the same as (3) but the 'sign penalty' $SP=3$ is used (4).

Table 1. *Particle parameters as functions of resolution*

	Lake's model			Yonath's model		
L	R_g (Å)	V (10^6 Å^3)	D_{max} (Å)	R_g (Å)	V (10^6 Å^3)	D_{max} (Å)
3	70·8	2·58	205	67·1	2·45	194
5	73·9	2·67	231	68·8	2·52	206
7	75·1	2·70	239	70·2	2·57	219

to a large extent in their contributions to the scattering curve (see Fig. 2). Minimization of such a functional seems to be an unrealistic task.

A realistic approach would be to increase the resolution step by step, that is, to fit first an inner portion of the SAS curve with lower multipoles, then to add higher harmonics taking into account a wider angular range *etc.* Since at each step the termination effects can lead to an overestimate of higher harmonics, the f_{lm}'s could be modified by a damping factor, *e.g.* $f_{lm}\rightarrow f_{lm}\exp[-\nu^2(l/L)^2]$ before the next step.

There is also a possibility of programming the expectation of termination effects into the functional. To do this, let us consider qualitatively the differences between 'truncated' and total intensities. Obviously, for higher angles the 'truncated' intensity is lower than the total one as a result of the missing higher multipoles. This is also valid for the scattering curve very close to the origin since the radius of gyration and in particular the maximum diameter increase markedly with improving resolution. The particle volume also increases with resolution, but not as much as D_{max} (see Table 1). One can therefore expect that in the central part of the scattering curve the 'truncated' intensity will be larger than the total one.

This behavior can be observed for the two models used (Fig. 6*a*). The s values where the sign of the difference changes correspond approximately to the maximum of the Porod plot (s_1) and the end of its steep slope (s_2), see Fig. 6(*b*). Model calculations with other bodies show similar features.

This can be used in the minimization process by introducing a 'sign' function

$$P(s)=\begin{cases}-1 & 0\le s<s_1-\delta s,\quad s\ge s_2+\delta s\\ 0 & s_1-\delta s\le s\le s_1+\delta s,\\ & s_2-\delta s\le s\le s_2+\delta s\\ +1 & s_1+\delta s<s<s_2-\delta s\end{cases}\qquad(20)$$

which describes the expected sign of the difference $I_{exp}(s)-I_{mod}(s)$. Here $\delta s>0$ is introduced to show that s_1 and s_2 are not known exactly but only estimated. If the sign of the difference does not coincide with $P(s)$ and $P(s)\neq 0$, the difference is to be multiplied by a factor $SP\ge 1$ ('sign penalty'). The value of SP can be taken as 2-3 at lower resolution and gradually decreased to unity with improving resolution.

The resulting curve obtained at $L=3$ for Lake's model with $SP=3$ is presented in Fig. 6 [curve (4)].

It is much closer to the 'truncated' curve. The figure of merit is $M_3 = 0{\cdot}066$ allowing further refinement.

The procedure using the 'sign penalty' function has been applied for the shape restoration of the two models. In both cases we started with a resolution $L = 3$ and progressively added higher harmonics. The restoration for Lake's model at $L = 5$ is compared to the initial model at the same resolution in Figs. 7(*a*), (*c*). The figure of merit is $M_5 = 0{\cdot}100$, whereas the M_3 value is only 0·047, indicating that most of the strongest contribution to the disagreement comes from the higher multipoles. Results of the restoration procedure for the SAS curve corresponding to Yonath's model are shown in Figs. 7(*b*), (*d*) ($M_3 = 0{\cdot}075$, $M_5 = 0{\cdot}147$). It is worth noting that all the minimization conditions were *exactly* the same for both cases. The shape restoration at this resolution was performed as a straightforward procedure, only the scattering curves were different (at $L = 5$ the range up to $s_{max} = 0{\cdot}08$ Å^{-1} is used). The minima of the functional are 4×10^{-3} and 7×10^{-3}, respectively.

Further refinement is possible but more ambiguous. Use of more multipoles yields a perfect fit to the scattering data, but not necessarily a better shape restoration. At the resolution $L = 7$ one can generate different shapes corresponding to the local minima of the functional by varying the damping factor, fixing or freeing lower multipoles *etc.* The possibility to choose the best model still exists, provided the scattering curve beyond the fitting region is available. An example is shown in Fig. 8. Here the scattering curves from two restorations of Lake's model are presented; curve (2) corresponds to the shape with $M_7 = 0{\cdot}125$, (3) to the one with $M_7 = 0{\cdot}104$. The fit in the range $0 < s < 0{\cdot}1$ Å^{-1} (used in the minimizing functional) is practically the same for the two curves (3×10^{-3} and 2×10^{-3}). It is, however, seen that curve(3), corresponding to the better figure of merit, fits the outer part of the scattering curve (*cf.* Fig. 4*a*) much better.

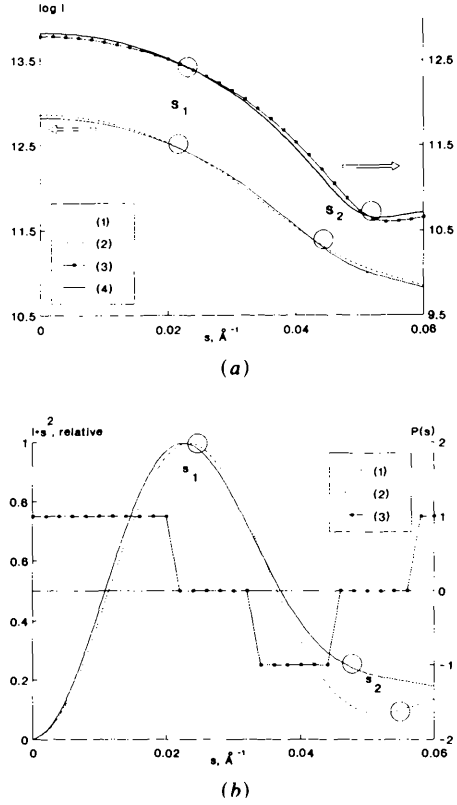


Fig. 6. (*a*) Truncated (1) and (3) and total (2) and (4) intensities from Lake's and Yonath's models, respectively; (*b*) Porod plots [total curves: (1) Lake, (2) Yonath] and the sign function (3). The points where the difference changes sign are marked.

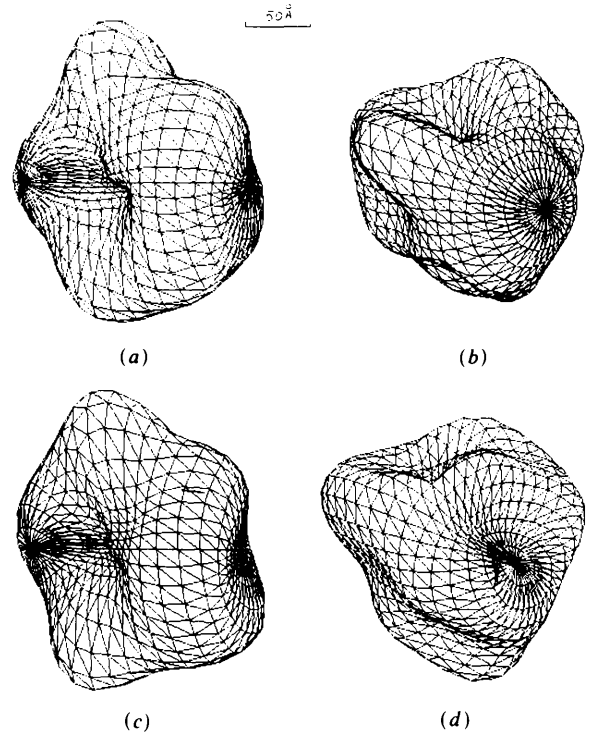


Fig. 7. Models of (*a*) Lake and (*b*) Yonath up to a resolution of $L = 5$ and their restored shapes, (*c*) and (*d*), respectively.

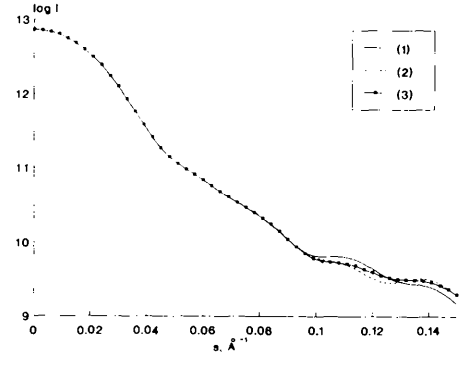


Fig. 8. Scattering curve from Lake's model (1) and two restored models with $L = 7$, $s_{max} = 0{\cdot}1$ Å^{-1}, (2) and (3).

The agreement of the scattering curve beyond the fit region can serve as a criterion to select the most plausible solution. The shape restorations thus obtained are presented in Fig. 9 (the figures of merit are 0·104 for Lake's model and 0·165 for Yonath's model). In both cases the angular range up to $s_{max} = 0{\cdot}1$ Å^{-1} was used for minimizing the functional and the range $0{\cdot}1 < s < 0{\cdot}15$ Å^{-1} for the choice of the 'best' solution.

Concluding remarks

The improvements presented above offer the possibility of more extensive practical use of the shape determination method. The calculations with model bodies enabled the development of a strategy for shape refinement allowing shape restoration with reasonable resolution.

Some limitations of the technique should, however, also be stressed. Firstly, the method, as well as the multipole expansion itself, is applied mostly to globular particles. The shape of such particles can always be described with (5) and the f_{lm} coefficients in (6) decrease rapidly with increasing index l. One can hardly expect successful results dealing with essentially anisometric objects. Secondly, the solution at higher resolution is by no means unique, especially if one takes into account the accuracy with which the shape scattering curves can be obtained in practice.

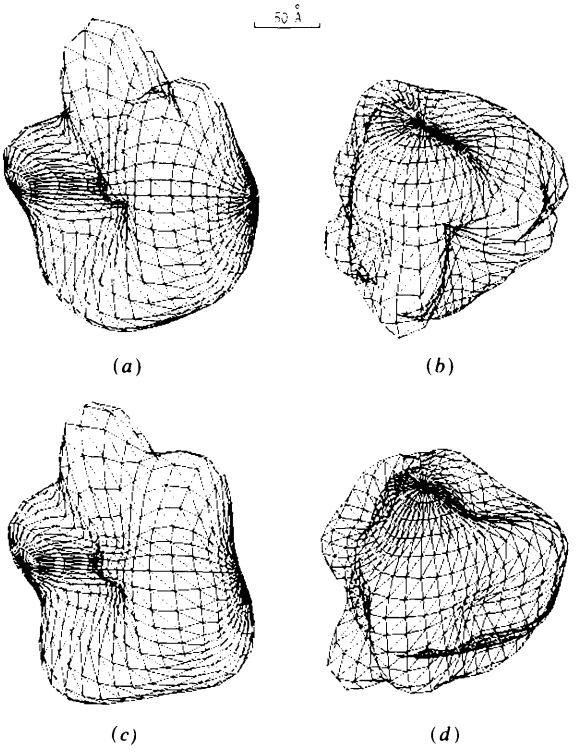


Fig. 9. Models of (*a*) Lake and (*b*) Yonath up to a resolution of $L = 7$ and their best restorations, (*c*) and (*d*), respectively.

The ability of the technique to fit the SAS curves neatly using a sufficiently large number of multipoles may become a shortcoming because the contribution from the internal particle structure may not be totally excluded from the experimentally obtained 'shape' scattering curves [for example, due to H-D exchange in neutron contrast variation experiments, see Stuhrmann (1975) or Witz (1983)]. Moreover, due to the sharp increase of the number of coefficients f_{lm} with the number of harmonics, $L = 6$–7 represents the upper limit of the resolution which one can hope to obtain, unless some additional information, for example about particle symmetry, is available. The refinement at higher resolution is also a time-consuming procedure: one refinement cycle at $L = 7$ requires approximately 12 h CPU time using a VAX 6000-410 computer.

In general, the shape determination with resolution up to $L = 3$–5 can be performed in a straightforward manner. The method is sufficiently sensitive to distinguish between the lower-order multipoles. Further refinement, however, is a difficult procedure. It is preferable not to attempt to fit the SAS curve in the complete available range of momentum transfer, but to use the tail portion of the curve for controlling the solution, even at the expense of resolution.

The authors are indebted to Professor A. Yonath for providing the cross sections of the model of the 50S ribosomal subunit and to Dr M. H. J. Koch for valuable comments on the manuscript. They are grateful to Drs V. Volkov and V. Bekhterev (Institute of Crystallography, USSR Academy of Sciences) for providing the original source code of the program package *OPTIS* and help in computer graphics, respectively. One of the authors (DS) is grateful for the support which he received from the GKSS Research Centre.

References

EDMONDS, A. R. (1957). *Angular Momentum in Quantum Mechanics.* Princeton Univ. Press.

FEIGIN, L. A. & SVERGUN, D. I. (1987). *Structure Analysis by Small-Angle X-ray and Neutron Scattering.* New York: Plenum Press.

GILL, P. E., MURRAY, W. & WRIGHT, M. H. (1981). *Practical Optimization.* London: Academic Press.

GLATTER, O. (1980). *Acta Phys. Austriaca*, **52**, 243-256.

GLATTER, O. & KRATKY, O. (1982). Editors. *Small-Angle X-ray Scattering.* London: Academic Press.

HARRISON, S. C. (1969). *J. Mol. Biol.* **42**, 457-483.

LAKE, J. (1976). *J. Mol. Biol.* **165**, 131-159.

MÜLLER, J. J. (1983). *J. Appl. Cryst.* **16**, 74-82.

MÜLLER, J. J., DAMASCHUN, G. & HÜBNER, G. (1979). *Acta Biol. Med. Ger.* **38**, 1-10.

ROLBIN, YU. A., KAYUSHINA, R. L., FEIGIN, L. A. & SCHEDRIN, B. M. (1973). *Kristallografiya*, **18**, 701-709. [Engl. transl: *Sov. Phys. Crystallogr.* (1974), **18**, 422-444.]

SHANNON, C. E. & WEAVER, W. (1949). *The Mathematical Theory of Communication.* Univ. of Illinois Press.

STUHRMANN, H. B. (1970*a*). *Acta Cryst.* A**26**, 297-306.
STUHRMANN, H. B. (1970*b*). *Z. Phys. Chem. (Frankfurt am Main)*, **72**, 177-184, 185-198.
STUHRMANN, H. B. (1975). *Small-Angle Scattering of Proteins in Solution.* In *Neutron Scattering for the Analysis of Biological Structures. Brookhaven Symp. Biol.* **27**, IV-3-19.
STUHRMANN, H. B. & KIRSTE, R. G. (1965). *Z. Phys. Chem. (Frankfurt am Main)*, **46**, 247-250.
STUHRMANN, H. B., KOCH, M. H. J., PARFAIT, R., HAAS, J., IBEL, K. & CRICHTON, R. R. (1977). *Proc. Natl Acad. Sci. USA*, **74**, 2316-2320.
TAUPIN, D. & LUZZATI, V. (1982). *J. Appl. Cryst.* **14**, 289-300.
WITZ, J. (1983). *Acta Cryst.* A**39**, 706-711.
YONATH, A. (1989). Personal communication.
YONATH, A., LEONARD, K. R. & WITTMANN, H. G. (1987). *Science*, **236**, 813-816.

Acta Cryst. (1991). A**47**, 744-748

E-Map Improvement in Direct Procedures

BY A. ALTOMARE, G. CASCARANO AND C. GIACOVAZZO

Istituto di Ricerca per lo Sviluppo delle Metodologie Cristallografiche, CNR, c/o Dipartimento Geomineralogico, Campus Universitario, 70124 *Bari, Italy*

AND D. VITERBO

Dipartimento di Chimica, Università della Calabria, 80030 *Arcavacata di Rende, Cosenza, Italy*

(*Received* 11 *October* 1990; *accepted* 3 *June* 1991)

Abstract

The quality of an E map is usually affected by systematic and/or random phase errors, by amplitude truncation effects in the series representation of the electron density, by the experimental uncertainty in the estimation of $|E|$ and by the intrinsic nature of the Fourier coefficients used (*i.e.* the E's). It is shown that simple supplemental calculations can improve the quality of an E map. Large molecular fragments can often be localized in the new map even when the original one is not easily interpretable.

1. Introduction

Usually a direct-methods procedure ends with one or more sets of approximated phases with which E maps are computed. The chemical significance of any trial solution is checked *via* atomic connectivity tables. Often one or more chemically sound fragments are well localized: the complete structure is then obtained by traditional least-squares and Fourier techniques. Sometimes a clear solution is not obtained: the map is uninterpretable because some atoms are occasionally missed or ghost peaks are present or the molecular geometry is distorted. But even in these cases a post mortem analysis of the structure may reveal the presence of correctly positioned atoms or fragments. Unfortunately, if such atoms or fragments were not *a priori* recognized in the E map, a procedure devoted to recovering the total from a partial structure (Karle, 1970; Beurskens, Prick, Doesburg, & Gould, 1979; Giacovazzo, 1983; Burla, Cascarano, Fares, Giacovazzo, Polidori & Spagna, 1989) would not be readily applicable. In this context it is of crucial importance to have a procedure which, from a traditional E map, is able to obtain a new more interpretable Fourier map.

The quality of a map depends on several factors among which the following three play a prominent role:

(*a*) *The phase errors.* In most cases these are unavoidable: large random errors can be tolerated without great loss of structural information in the E map while systematic errors have greater destructive effects (Silva & Viterbo, 1980).

(*b*) *Amplitude truncation effects* in the series representation of the electron density. Traditional direct methods do not phase reflections under the minimum threshold value $E_{\mathrm{Tr}} \simeq 1{\cdot}2$: in most of the practical applications E_{Tr} lies in the range 1·30–1·50. If phases are determined with sufficient accuracy the amplitude truncation effects are not really harmful (this is a necessary condition for the general success of direct methods). However, if this effect is associated with phase errors the final result is often destructive. A classical example is structures suffering from pseudo-translational symmetry: if no special action is undertaken the reflections actively used in the phasing process coincide with substructure reflections. Even when these reflections are accurately phased, and that it is not the rule, the information on the superstructure is completely lost in the E map.

(*c*) *The Fourier coefficients* used for calculating the map. It is a traditional practice to use E coefficients at the conclusion of a phasing process: they produce

0108-7673/91/060744-05$03.00