Error Analysis of a Smeared SAXS Curve of Polystyrene in Benzene*

BY O. KUBE AND J. SPRINGER

Institut für Technische Chemie der Technischen Universität Berlin, Fachgebiet Makromolekulare Chemie, Strasse des 17. Juni 135, 1000 Berlin 12, Federal Republic of Germany

(Received 8 August 1984; accepted 26 April 1985)

Dedicated to Professor Dr R. Bonart on his 60th birthday

Abstract

An error analysis is performed for a slit-length-smeared SAXS curve of polystyrene in benzene. The desmearing and further calculations are done by indirect Fourier transformation. Four different functions are considered. The results are compared and the different error propagation properties of the chosen functions are discussed.

Introduction

Randomly oriented macromolecules in a solution can be described by the averaged one-dimensional distance distribution function \( p(r) \). The normalized scattered intensity \( I(h) \) [\( I(h) \) is dimensionless but for convenience will be further referred to as an intensity] and the distance distribution function are connected by an integral transformation (Glatter & Kratky, 1982, pp. 124, 130):

\[
I(h) = 4\pi \int_0^\infty p(r) \sin(hr)/hr \, dr 
\]

\[
p(r) = 1/(2\pi^2) \int_0^\infty \tilde{I}(h) hr \sin(hr) \, dr. 
\]

\( h = 4\pi/\lambda \sin\theta \), \( \lambda \): wavelength of the radiation, \( 2\theta \): angle between primary beam and scattered radiation, \( r \): distance between scattering centers.

The usually measured mean-square radius of gyration \( \langle R^2 \rangle \) \( (R_{gyr} = \langle R^2 \rangle^{1/2}) \) of a particle can be calculated from the distance distribution function

\[
\langle R^2 \rangle = \int_0^\infty r^2 p(r) \, dr \left[ 2 \int_0^\infty p(r) \, dr \right].
\]

The zero intensity (used to determine the molar weight is connected to the distance distribution function according to

\[
I(0) = 4\pi \int_0^\infty p(r) \, dr. 
\]

Equations (1) and (2) are strictly valid only for point collimation and monochromatic radiation. The experimentally observed small-angle scattering is usually smeared owing to the finite geometry of the primary beam and the wavelength distribution of the radiation used. For monotonically decaying scattering curves it is sufficient to correct the smearing effect of the finite length of the profile of the primary beam \( P(t) \). The smeared intensity \( \tilde{I}(h) \) is then related to the desmeared intensity \( I(h) \) by (Glatter & Kratky, 1982, pp. 121–122)

\[
\tilde{I}(h) = \int_0^{t_{max}} P(t) I\left( [h^2 + t^2]^{1/2} \right) dt,
\]

\( t \) is the coordinate along the length of the primary beam, \( P(t) \) is symmetric to \( t = 0 \) nm\(^{-1}\) and normalized to

\[
\int_0^{t_{max}} P(t) \, dt = 1.
\]


The usual way to process the scattering curve is to desmear it and to determine the radius of gyration and zero intensity via a Guinier or Zimm plot. This procedure will give consistent results whenever the zero slope of the scattering curve is nearly approached.

When it is not possible to restrict experiments to small particles or to adjust the resolution accordingly the above-mentioned procedure will give ambiguous results owing to the lower precision of the plots and a systematic error introduced by the desmearing routine. In that case it is advisable to use all the information from the scattering curve and to check for introduced systematic errors. This can conveniently be done with the indirect Fourier transformation method (Glatter, 1977a, b; Hossfeld, 1968; Moore, 1980), where different approximating functions for the distance distribution function are tested. The unknown parameters of these functions are calculated by a least-squares fit of the transformed approximating function to the smeared data. The desmeared intensity will be a by-product.

*Presented partly at the spring meeting of the Deutsche Physikalische Gesellschaft, Münster, 12–16 March 1984.
The minimum requirements that must be met by the experiment are stated by the Fourier sampling theorem (Bracewell, 1978):
1. there must exist a maximum correlation length \( r_{\text{max}} \) for which \( p(r > r_{\text{max}}) = 0 \) (band-limited function);
2. the intensity must be measured up to \( h_{\text{max}} = \pi/r_{\text{max}} \);
3. the distance \( \Delta h \) between measured points must be \( \Delta h = \pi/r_{\text{max}} \).

The additional requirement in the case of continuous small-angle scattering that the curve should be measured up to an angle for which \( I(h) = 0 \) usually cannot be met for scattering curves of polymers in solution.

**Experimental**

Measurements were done with a Kratky camera (Anton Paar, Graz, Austria, model 11000). The 0.06 mm entrance slit and the 14 mm slit-length limiting stop were used. The X-ray generator was operated with a Cu X-ray tube of focal width 1 mm at 40 mA and 40 kV. The distance between sample and position-sensitive detector (M. Braun, Garching, Federal Republic of Germany, model PSD 50) was set to 281 mm. A single channel of the PSD 50 refers to an area of 0.042 x 5 mm².

The slit-length profile \( P(t) \) was determined with a slit-length measuring device (Anton Paar, model 12300) in the sample plane (Fig. 1). The original slit-length profile is approximated with a cubic spline function (equation 6), which is folded with a 5 mm step function. The halves of the profile are averaged and the resulting half is interpolated again at the fixed \( J \) knots \( t_j \) with a cubic spline function and normalized (Fig. 1):

\[
P(t) = \sum_{j=1}^{J-1} \sum_{l=0}^{3} s_{ij}t^l \quad \text{with} \quad s_{ij} = \begin{cases} 0 & t < t_j \\ s_{ij} & t_j \leq t < t_{j+1} \\ 0 & t \geq t_{j+1} \end{cases}
\]

This procedure has the advantage that in most cases the smearing operation can be carried out analytically using the cubic spline representation of the slit-length profile.

The scattering of the solvent benzene and polystyrene in benzene (PS/BE, \( M = 110,000 \text{ g mol}^{-1}, T = 293 \text{ K} \)) were measured four times, 0.5 h each, averaged and subtracted. The number of points of the original scattering curve recorded were numerically reduced by combining up to 11 channels.

**Indirect Fourier transformation method**

The indirect Fourier transformation developed by Glatter (1977a,b, program ITP79) approximates the distance distribution function by a sum of cubic \( B \) splines \( B_k \) up to \( r_{\text{max}} \):

\[
p(r) = \sum_{k=1}^{N} c_k B_k.
\]

The \( B \) splines are transformed according to (1) and (5) and the linear coefficients \( c_k \) are determined with a least-squares fit. To represent the distance distribution function sufficiently many cubic \( B \) splines are necessary and so the fit is constrained to avoid unwanted fluctuations.

In the case considered the Lagrange multiplier is found at the very beginning of the plateau of the stability plot. The distance distribution function obtained is very sensitive to small changes of the Lagrange multiplier.

The calculated results show a strong dependence on \( r_{\text{max}} \). Since \( r_{\text{max}} \) is used as a fixed (correct) input parameter the influence on the propagated errors is neglected. Several values for \( r_{\text{max}} \) must be tested. A good starting value is obtainable from the fit with Gaussian functions. The number of \( B \) splines has only a minor influence on the fit (usually 30 cubic \( B \) splines are sufficient).

The correlation function obtained is shown in Fig. 5. The fact that the function does not reach 0 at \( r = 0 \text{ nm} \) (constant background in \( h \) space) is a result of the cut-off \( B \) splines at the origin and the width of the equidistant \( B \) splines, which cannot be made smaller than \( r_{\text{max}}/40 \). If the knots for the \( B \) splines could be optimized a function that fits the data and reaches 0 precisely at \( r = 0 \text{ nm} \) could be constructed. Even though the last statement is not proved Figs. 6 and 7 provide strong arguments (see also Debye fit).

**Error propagation**

The error propagation (see Appendix) for zero intensity [see (4)] and radius of gyration [see (3)] can easily be added to program ITP79. The integration limits for \( B \) splines correspond to their range of definition.
Zero intensity $I(0)$:

$$V_k = 4\pi \int B_k \, dr$$

$$\sigma^2_{I(0)} = V^T CV.$$  

Radius of gyration $R_{gyr}$:

$$V_k = \pi/[I(0)R_{gyr}] \left( \int B_k r^2 \, dr - 2 \langle R^2 \rangle \int B_k \, dr \right)$$

$$\sigma^2_{R_{gyr}} = V^T CV.$$  

**Approximation with Gaussian functions**

The smeared intensity $\tilde{I}(h)$ can be approximated by Gaussian functions with the center at $h = 0$ nm$^{-1}$.

$$\tilde{I}(h) = \sum_{k=1}^{N} a_{2k-1} \exp(-a_{2k} h^2).$$  \hspace{1cm} (8)

The desmeared intensity $I(h)$ is given by

$$I(h) = \sum_{k=1}^{N} a_{2k-1} M_{02k-1} \exp(-a_{2k} h^2)$$  \hspace{1cm} (9)

with

$$M_{02k-1} = \int_0^{t_{\text{max}}} P(t) \exp(-a_{2k} t^2) \, dt.$$  \hspace{1cm} (10)

[This approach is chosen to avoid the calculation of the integral in (10) during each iteration step of the program.] The integration in (10) can be carried out analytically inserting (6) for the slit-length profile $P(t)$ into (10). The integration must be carried out stepwise for each spline interval from $t_j$ to $t_{j+1}$. The resulting integrals with odd ($O$) and even ($E$) powers of $t$ are conveniently solved via recursion schemes.

$$O_1 = -\exp(-a_{2k} t_j^2)/(2a_{2k}) \bigg|_{t_j}^{t_{j+1}}.$$  

$$O_{2l+1} = \left[ t^{2l} \exp(-a_{2k} t^2) \bigg|_{t_j}^{t_{j+1}} - lO_{2l-1} \right]/(2a_{2k}).$$  

$E_0$ can be calculated by means of the Fortran supplied error function erfc.

$$\text{erfc}(x) = 2x^{-1/2} \int_{x}^{\infty} \exp(-z^2) \, dz$$  \hspace{1cm} (11)

$$E_0 = 0.5(\pi/a_{2k})^{0.5}[\text{erf}(a_{2k}^{-0.5}t_j) - \text{erf}(a_{2k}^{-0.5}t_{j+1})].$$  

For $l = 1, 2, ...$

$$E_{2l} = -0.5/a_{2k} \left[ t^{2l-1} \exp(-a_{2k} t^2) \bigg|_{t_j}^{t_{j+1}} + 1/(2l-1)E_{2l-1} \right].$$

The Gaussian functions can be transformed analyti-}

$$p(r) = \sum_{k=1}^{N} \frac{a_{2k-1} M_{02k-1}}{(4\pi a_{2k})^{1.5}} r^2 \exp(-r^2/(4a_{2k})).$$  \hspace{1cm} (12)

The mean-square radius of gyration can be calculated according to (3):

$$\langle R^2 \rangle = 3 \sum_{k=1}^{N} a_{2k-1} M_{02k-1} a_{2k}$$

$$\times \left[ \sum_{k=1}^{N} a_{2k-1} M_{02k-1} \right]^{-1}.$$  \hspace{1cm} (13)

The fit can be done without constraints over the whole measured range using a gradient method (Brandt, 1981; Stoer, 1979). The fit converges rapidly against a minimum and is rather independent of initially chosen values for the coefficients. The number of Gaussian functions (usually four) is taken to be appropriate if the least-squares norm of fit [LSQN, see also (33)] is smaller than the number of measured points $M$ and if the resolution of the distance distribution function is sufficient. The smeared fit is shown in Fig. 2.

The distance distribution function (Fig. 6) indicates a maximum correlation length $r_{\text{max}} \approx 18$ nm. This value can be used immediately for the indirect Fourier transformation method (program ITP79).

The tail of the curve measured is fitted by a slowly decaying Gaussian function, which is visible as a small spike at $r = 0.3$ nm in the distance distribution function (Fig. 6). This is equivalent to the 'background' observed with the indirect Fourier transformation method.

**Error propagation**

The covariance matrix $C$ (see Appendix) is obtainable from the last iteration step where the new coefficients are calculated from the linearized function. The covariance matrix therefore contains the variances for the coefficients $a_1$ to $a_{2N}$.

![Fig. 2. Fit with GAUSS1 to the smeared data.](image-url)
In the following equations \( N \) Gaussian functions are considered, i.e. \( k \) runs from 1 to \( N \), vector \( V \) has \( 2N \) components and \( C \) is a \( 2N \) square matrix.

Smeared zero intensity \( \bar{I}(0) \) [see (8)]:
\[
V_{2k-1} = 1 \quad \text{and} \quad V_{2k} = 0.
\]

Desmeared intensity \( I(h) \) [see (9)]: \( M_{02k-1} \) (10) is a function of \( a_{2k} \) and yields differentiated
\[
M_{2k-1} = dM_{02k-1}/da_{2k} \\
= - M_{02k-1} t_{\text{max}} \int_0^{t_{\text{max}}} P(t) t^2 \exp(-a_{2k} t^2) \, dt.
\]

The integral in (14) can be solved with the same scheme as proposed for the integral in (10).
\[
V_{2k-1} = M_{02k-1} \exp(-a_{2k} h^2), \\
V_{2k} = a_{2k-1} \exp(-a_{2k} h^2) (M_{22k-1} - h^2 M_{02k-1}).
\]

Desmeared zero intensity \( I(0) \) [see (9)]:
\[
V_{2k-1} = M_{02k-1}, \\
V_{2k} = a_{2k-1} M_{22k-1}.
\]

Radius of gyration \( R_{\text{gyr}} \) \( [R_{\text{gyr}} = \langle R^2 \rangle^{1/2}, \text{see (13)}] \):
\[
V_{2k-1} = 3/[2 R_{\text{gyr}} I(0)] M_{02k-1} (a_{2k} - \langle R^2 \rangle/3), \\
V_{2k} = 3/[2 R_{\text{gyr}} I(0)] a_{2k-1} \times [M_{02k} + M_{22k-1}(a_{2k} - \langle R^2 \rangle/3)].
\]

Distance distribution function \( p(r) \) [see (12)]:
\[
V_{2k-1} = r^2(4\pi a_{2k})^{-1-5} M_{02k-1} \exp(-r^2/4a_{2k}), \\
V_{2k} = r^2 a_{2k-1}(4\pi a_{2k})^{-1-5} \exp(-r^2/4a_{2k}) \times [M_{02k-1}(r^2/4a_{2k} - 1.5/a_{2k}) + M_{22k-1}].
\]

Debye fit

A macromolecule can be approximated by a model of \( N + 1 \) connected segments. Assuming a Gaussian distance distribution between two arbitrary segments \( i \) and \( j \) \((i \neq j)\), one arrives at the scattering function \( D(h, R_{\text{gyr}}) \) of a single macromolecule, which was calculated by Debye (1947):
\[
D(h, R_{\text{gyr}}) = 2u^{-2}(e^{-u} + u - 1), \quad u = h^2 \langle R^2 \rangle. \quad (21)
\]

The theory has the drawbacks that a Gaussian distance distribution is only a valid approximation for widely separated segments under \( \theta \) conditions and that the internal structure of the polymer chain is neglected. Nevertheless, a satisfactory fit is obtainable for low concentrations \((c < 10 \text{ g}^{-1})\) in a restricted \( h \) range \( h R_{\text{gyr}} < 10 \). To extend the fit to larger \( h \) values and to higher concentrations additional degrees of freedom are necessary. As a first approximation a constant background was determined assuming an \( h^{-3} \) course for the tail of the scattering curve.
\[
\bar{I}(h) = \text{constant} \times h^{-3} + \text{background}.
\]

The concentration dependence of this background shows that it is an inherent element of the scattering curve, most probably unresolved scattering of the subunits of the polymer chain.

The constant background is a useful numerical approximation but untractable for further analytical treatment. A much better suited function with the same characteristics is a slowly decaying Gaussian function. Therefore, the smeared intensity is approximated according to (12). \([a_2 \text{ is substituted in (22) for } R_{\text{gyr}} \text{ in (21) because the further use of } R_{\text{gyr}} \text{ at this point would be misleading.}]\)
\[
\bar{I}(h) = a_1 \bar{D}(h, a_2) + a_3 \exp(-a_4 h^2) \quad (22)
\]
\[
\bar{D}(h, a_2) = \frac{t_{\text{max}}}{0} P(t) D([h^2 + t^2]^{1/2}, a_2) \, dt.
\]

For non-linear least-squares methods the derivative \( d\bar{D}(h, a_2)/da_2 \) is needed. It can be calculated without approximation.
\[
d\bar{D}(h, a_2)/da_2 = \frac{t_{\text{max}}}{0} P(t) \{dD([h^2 + t^2]^{1/2}, a_2)/da_2\} \, dt \quad (23)
\]
\[
dD(h, a_2)/da_2 = 2/a_2 [2 - D(h, a_2)(h^2a_2^2 + 2)].
\]

The desmeared intensity is given by \([\text{see (9) with } k = 2]\)
\[
I(h) = a_1 D(h, a_2) + a_3 M_{02} \exp(-a_4 h^2). \quad (24)
\]

From the model of \( N + 1 \) segments the distance distribution function \( p_D(r) \) for the Debye function \( D(h, a_2) \) can be calculated \([\text{erfc: error function, see (11)}]\):
\[
p_D(r) = (r^2/a_2^2 + r^2/4a_2^2)/\pi \text{erfc}(r/2a_2) \\
- r^2/\pi a_2^2 \exp[-(r/2a_2)^2]. \quad (25)
\]

The distance distribution function for one Gaussian function \([\text{with } k = 2]\) is already known and so the complete distance distribution function \( p(r) \) is given by
\[
p(r) = a_1 p_D(r) + a_3 M_{02}/(4\pi a_2^4)^{1.5} r^2 \exp(-r^2/4a_2^4). \quad (26)
\]

The mean-square radius of gyration is simply given by
\[
\langle R^2 \rangle = (a_1 a_2^3 + 3a_3 M_{03} a_2)/(a_1 + a_3 M_{03}). \quad (27)
\]

\textbf{Error propagation}

The covariance matrix \( C \) (see Appendix) as for the fit with Gaussian functions is obtained from the last
iteration step where the coefficients for the linearized function are calculated. \( C \) contains covariances of \( a_1 \) to \( a_4 \).

Smeared zero intensity \( \tilde{I}(0) \) [see (22)]:
\[
V_1 = \tilde{D}(0, a_2), \quad V_2 = a_1 \frac{d\tilde{D}(0, a_2)}{da_2} \quad \text{[see (23)]},
V_3 = 1, \quad V_4 = 0.
\]

Desmeared intensity \( I(h) \) [see (24)]:
\[
V_1 = D(h, a_2), \quad V_2 = a_1 \frac{dD(h, a_2)}{da_2} \quad \text{[see (23)]},
V_3\text{ and } V_4\text{ are the same as in (15) and (16) with } k = 2.
\]

Desmeared zero intensity \( I(0) \) [see (24)]:
\[
V_1 = 1, \quad V_2 = 0,
V_3\text{ and } V_4\text{ are the same as in (17) and (18) with } k = 2.
\]

Radius of gyration \( R_{\text{gyr}} \) [see (27)]:
\[
V_1 = \frac{1}{2R_{\text{gyr}}(0)} (a_2^2 - \langle R^2 \rangle),
V_2 = \frac{1}{2R_{\text{gyr}}(0)} 2a_1 a_2,
V_3 = M_03/\{2R_{\text{gyr}}(0)[3a_4 - \langle R^2 \rangle],
V_4 = 1/[2R_{\text{gyr}}(0)][3a_3 M_03 + M_23(\langle R^2 \rangle - 3a_3 a_4)].
\]

Distance distribution function \( p(r) \) [see (25) and (26)]:
\[
V_1 = p_0(r),
V_2 = a_1/\{\pi a_2^2\} [-r/a_2\{1 + (r/a_2)^2\} \text{erfc}(r/2/a_2)
+ 2/\pi^{0.5}(r/a_2)\exp[-(r/2/a_2)^2]},
V_3\text{ and } V_4\text{ are the same as in (19) and (20) with } k = 2.
\]

Polynomial fit
For curves that decay monotonically such as the measured ones a polynomial fit can generally be tried.
\[
\tilde{I}(h) = \sum_{k=0}^{2N+1} a_k h^k. \quad \text{(28)}
\]
To obtain the smeared \( \tilde{h}^k \) it is advantageous to treat even \( (2k, k = 0, \ldots, N) \) and odd \( (2k+1, k = 0, \ldots, N) \) parts separately. With the knowledge of the moments \( \text{Mom}_{2l} \) (29) of the length distribution of the primary beam the even parts of the polynomials can easily be smeared analytically. [For the purpose of error propagation the smearing approach is superior to the desmearing approach devised by Walter & Schmidt (1981).]
\[
R_{2k} = \sum_{l=0}^{k} \binom{k}{l} h^{2k-2l} \text{Mom}_{2l}
\]
with
\[
\text{Mom}_l = \int_0^{t_{\text{max}}} P(t) t^l \, dt \quad \text{(29)}
\]
\[
h_{2k+1} = \int_0^{t_{\text{max}}} P(t) (h^2 + t^2)^{k+1}/(h^2 + t^2)^{0.5} \, dt. \quad \text{(30)}
\]

Inserting (6) for the slit-length profile \( P(t) \) in (30) the smearing can be carried out analytically. The resulting integrals of the type
\[
\int_{t_j}^{t_{j+1}} t^l(h^2 + t^2)^{k+1}/(h^2 + t^2)^{0.5} \, dt
\]
can easily be solved with the substitution \( z = h^2 + t^2 \) in the case \( l = 1, 3 \). For \( l = 0, 2 \) multiplication is carried out, leaving only elementary forms \( E[2m = 2(i+1) + l, i = 0, \ldots, k] \)
\[
E_{2m} = \int_{t_j}^{t_{j+1}} t^{2m}/(h^2 + t^2)^{0.5} \, dt,
\]
which in turn can successively be calculated using recursion:
\[
E_0 = \log_t\left[ t + (h^2 + t^2)^{0.5} \right]_{t_j}^{t_{j+1}}.
\]
For \( m = 1, 2, \ldots, \)
\[
E_{2m} = 1/2m \left[ t^{2m-1}(h^2 + t^2)^{0.5} \right]_{t_j}^{t_{j+1}}
- (2m - 1) h^2 E_{2(m-1)}.
\]
The desmeared fit is simply given by
\[
I(h) = \sum_{k=0}^{2N+1} a_k h^k. \quad \text{(31)}
\]
The distance distribution function can in principle be calculated analytically:
\[
p(r) = \sum_{k=0}^{2N+1} a_k \cos(hr)
\times \sum_{l=0}^{L} (-1)^{l+1} l!/(k - 2l)! h^{k-2l}/r^{2l}
+ \sin(hr) \sum_{m=0}^{M} (-1)^{m} k!/(k - 2m - 1)
\times h^{k-2m-1}/r^{2m-1} \int_{r=0}^{r_{\text{max}}} \, dt \quad \text{(32)}
\]
\( L = \text{Trunc}[(k + 1)/2] \) and \( M = \text{Trunc}(k/2) \). \text{Trunc} means that, for instance, \( \text{Trunc}(3.5) = 3 \).
Since the fit cannot be extended to larger values than \( h_{\text{max}} \) one has to deal with a significant termination effect. To restore a reasonable distance distribution function the values according to (32) are calculated at \( r = n\pi h_{\text{max}} / n = 1, 2, \ldots \) and then smoothed using a cubic \( B \) spline. The smoothed distance distribution function is accepted when deviations of the transformed smoothed distance distribution function from the calculated desmeared intensity are negligible. The transform of the distance distribution function in Fig. 8 differs less than 1% at small \( h \) values and less than 15% at large \( h \) values from the original desmeared polynomial in the measured range. As the normalization \( 4\pi \int p(r) \, dr = I(0) \)
Table 1. Polystyrene/benzene \( M = 110000 \) g mol\(^{-1} \), \( c = 30.0400 \) g l\(^{-1} \), \( T = 293 \) K

LSQVAR: Variance of the fit to \( M \) points with \( N \) parameters.
LSQVAR \( = \frac{1}{(M-N)\text{LSQN}} \) [see also (33)].

### Averaged distance distribution function:
\[
\bar{I}(0) = 1.144 \text{ counts s}^{-1}, \quad I(0) = 4.28 \text{ counts s}^{-1}, \quad R_{\text{gyr}} = 4.52 \text{ nm}.
\]

<table>
<thead>
<tr>
<th>Program</th>
<th>LSQVAR ( \bar{I}(0) ) (counts s(^{-1} ))</th>
<th>( I(0) ) (counts s(^{-1} ))</th>
<th>( R_{\text{gyr}} ) (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITPE79</td>
<td>0.545</td>
<td>1.40 \pm 7.3%</td>
<td>4.52 \pm 19.1%</td>
</tr>
<tr>
<td>GAUSS1</td>
<td>0.419</td>
<td>1.130 \pm 1.1%</td>
<td>4.14 \pm 3.5%</td>
</tr>
<tr>
<td>DEBYE3</td>
<td>0.546</td>
<td>1.104 \pm 0.9%</td>
<td>3.81 \pm 1.7%</td>
</tr>
<tr>
<td>POLFIT3</td>
<td>0.430</td>
<td>1.178 \pm 0.9%</td>
<td>4.94 \pm 2.5%</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td>1.137 \pm 3%</td>
<td>4.70 \pm 27%</td>
</tr>
</tbody>
</table>

*Truncated distance distribution function.

The fit (28) can be done over the whole measured range and extrapolated to \( h = 0 \) nm\(^{-1} \) without problems. Depending on the accuracy of the data, oscillations will occur in the range of large \( h \) values. The number of coefficients are taken to be appropriate if the least-squares norm LSQN is smaller than the number of measured points and the resolution of the distance distribution function is sufficient.

### Error propagation

The vector \( V \) has \( 2N + 2 \) components and the \( 2N + 2 \) square matrix \( C \) contains the variances for the coefficients \( a_k \) \((k = 0, \ldots, 2N + 1)\).

Smeared zero intensity \( \bar{I}(0) \) [see (28), (29)]:
\[
V_k = \text{Mom}_k.
\]

Desmeared intensity \( I(h) \) [see (31)]:
\[
V_k = h^k.
\]

Desmeared zero intensity \( I(0) \) [see (31)]:
\[
\sigma_{I(0)}^2 = \sigma_{a_0}^2.
\]

Distance distribution function \( p(r) \):

An error propagation can only be done for the original transformed polynomial. Assuming identical relative errors for the original \( p(r)_p \) and smoothed distance distribution function \( p(r)_s \), we can calculate a scaled variance for the smoothed function.

\[
V_k = \cos(hr) \sum_{l=0}^{L} (-1)^{l+1} \frac{l!(k - 2l)!}{(k - 2l)!} h^{k-2l} r^{2l} + \sin(hr) \sum_{m=0}^{M} (-1)^m k!(k - 2m - 1) \times h^{k-2m-1} r^{2m+1} \\
\times \frac{k_{\text{max}}^{h_{\text{max}}}}{h_0}, \quad L = \text{Trunc}[(k + 1)/2], \quad M = \text{Trunc}(k/2), \quad \sigma_{p(r)_s}^2 = [p(r)_s/p(r)_p]^2 V^TCV.
\]

### Discussion

The SAXS curve was measured with a reasonable resolution of \( h_{\text{max}} = 0.073 \) nm\(^{-1} \). The first 50 points are given with the original increment of \( \Delta h = 0.006 \) nm\(^{-1} \). This would allow full restoration of the distance distribution function up to \( \pi/h_{\text{max}} = 43 \) nm.

The smeared fits (Fig. 3) are all excellent approximations to the data. The fit with Gaussian functions is shown as a reference in Fig. 2.

The desmeared fits (Fig. 4) are nearly identical in the measured range but diverge in the extreme small-angle region. Since this part of the scattering curves is used for a Guinier or Zimm plot full advantage is taken of the possible deviations. As was expected the propagated errors for \( I(0) \) are larger than for \( \bar{I}(0) \) (see Table 1).

Fig. 5 shows the error band for the distance distribution function calculated with the indirect Fourier transformation method. The \( B \) splines at large \( r \) have their main feature close to the origin when Fourier transformed (see Glatter & Kratky, 1982, p. 134). The variances of the respective coefficients will mainly represent the statistics of the data close to \( h = 0 \) nm\(^{-1} \). The propagated errors for \( I(0) \) and \( R_{\text{gyr}} \) therefore cover
the observed real deviations (see Table 1). On the other hand, the need for stabilization is caused by the nearly linear dependence of these transformed $B$ splines.

The propagated errors obtained by the fit of Gaussian functions and the Debye fit are strikingly small compared to the observed real deviations (Table 1, Figs. 6 and 7). This is readily explained by the fact that only a few functions with broad extension are necessary to approximate the data (the narrowest Gaussian function used drops to zero at $h = 0.8$ nm$^{-1}$). The small errors for the fit parameters favor the use of these functions in the fitted range but give rise to far too low error estimates for quantities depending on the extrapolation to $h = 0$ nm$^{-1}$ such as $I(0)$ and $R_{gyr}$.

The differing results of the polynomial fit (Table 1) are caused by the fact that, contrary to all other functions, $[dI(h)/dh]_{h=0} \neq 0$. The structure of the error band of the distance distribution function (Fig. 8) is mainly caused by the heavy termination effect, which becomes smaller with larger $r$. Of course results calculated only with the polynomial fit would be of limited value but as an addition to better suited functions they are a great help to check for the limits of possible deviations allowed by the data. For comparison all calculated distance distribution functions are shown together in Fig. 9.

The differences at $r < 1$ nm result from the implicitly performed extrapolations to large $h$. These extrapolations are hardly covered by data and therefore
cannot seriously be considered for an interpretation. The deviations at \( r > 8 \text{ nm} \) are caused by the different extrapolations to \( h = 0 \text{ nm}^{-1} \). The indirect Fourier transformation method, the Gauss fit and the Debye fit explicitly or implicitly specify a band-limited distance distribution function and thereby consistently indicate a value of \( r_{\text{max}} \approx 18 \text{ nm} \). The polynomial fit on the other hand does not imply such a restriction and its distance distribution function might be taken as a possible upper limit at large \( r \) (see error band Fig. 8) allowed by the data. Judged from this point of view \( r_{\text{max}} \) is not an objective parameter but an optimized quantity depending on the used function, range and statistic of the data. The interaction of the statistic of the data and the determination of \( r_{\text{max}} \) was investigated by Gerber & Schmidt (1983).

To close the whole procedure the four distance distribution functions in Fig. 9 can be averaged (neglecting variances), transformed, smeared and again fitted (one parameter only) to the data. The fit obtained is again in very good agreement with the data.

It is well known that errors are magnified by desmearing (Strobl, 1970) but to the best of our knowledge the possible magnitude has never been intensively studied for real SAXS curves. The revealed error of 27% for the radius of gyration for the presented SAXS curve is surprisingly high. Of course it will be impractical to use more than one method for data processing in standard application but if high precision for the interpretation is necessary it will be advisable to compare results of different methods. This will generally hold when the resolution of the experiment approaches the maximum correlation length.

This work was sponsored by the Deutsche Forschungsgemeinschaft and the Fonds der Chemischen Industrie to whom we express our grateful thanks.

**APPENDIX**

Let \( y(h) \) denote a measured quantity with variance \( \sigma_i \) at \( h_i \) and \( F(h) \) the function to be fitted.

\[
F(h) = \sum_{j=1}^{N} a_j f_j(h_i).
\]

The \( f_j(h_i) \) are known predefined basic functions. The coefficients \( a_j \) are to be determined so that the weighted least-squares norm \( \text{LSQN} \) for \( M \) measured points is minimized.

\[
\text{LSQN} = \sum_{i=1}^{M} \left[ y(h_i) - F(h_i) \right]^2 / \sigma_i^2.
\]  

(33)

\( N \) sets of equations \( d\text{LSQN}/da_j \) \( (j=1,...,N) \) are obtained, which can be written in matrix notation as \( (F^T F) A = F^T Y \).

The elements of matrix \( F \) and vectors \( A \) and \( Y \) are

\[
F_{ij} = f_j(h_i) / \sigma_i, \quad A_j = a_j, \quad Y_i = y(h_i) / \sigma_i.
\]

The covariance matrix \( C \) (Brandt, 1981) contains the covariances of the fitted coefficients.

\[
C = (F^T F)^{-1}
\]

\[
C_{jk} = \sigma_j \sigma_k.
\]

The variance of any quantity \( X \) obtainable from the fitted function can be calculated by

\[
\sigma_X^2 = V^T CV,
\]

where the \( k \)th element of the column vector \( V \) is \( V_k = dX / da_k \).

**References**


