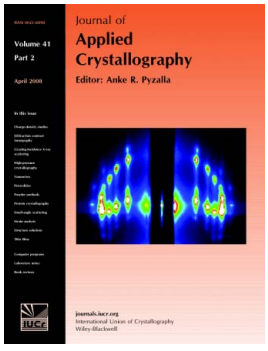
Journal of
Applied Crystallography
ISSN 0021-8898
Editor: **Anke R. Pyzalla**

Simultaneous estimation of the form factor and structure factor for globular particles in small-angle scattering

Steen Hansen

J. Appl. Cryst. (2008). **41**, 436–445



Journal of
Applied Crystallography
ISSN 0021-8898

Received 22 October 2007
Accepted 20 February 2008

Simultaneous estimation of the form factor and structure factor for globular particles in small-angle scattering

Steen Hansen

Department of Natural Sciences, Faculty of Life Sciences, University of Copenhagen, Thorvaldsensvej 40, DK-1871 FRB C, Denmark. Correspondence e-mail: slh@life.ku.dk

Small-angle scattering data from non-dilute solutions of particles are often analysed by indirect Fourier transformation using a specific model structure factor to obtain an estimate of the distance distribution function that is free from concentration effects. A new approach is suggested here, whereby the concentration effects are expressed solely through real space functions without the use of an explicit structure factor. This is done by dividing the total distance distribution function for the scattering into three different contributions, as suggested by Kruglov [(2005). *J. Appl. Cryst.* **38**, 716–720]: (i) the single particle distribution which is due to intraparticle effects, (ii) the excluded volume distribution from excluded volume effects which is only dependent upon the geometry of the particles, and (iii) a structure distribution which is due to the remaining interaction between the particles. Only the single particle distribution and the structure distribution are allowed to vary freely (within the restrictions of a smoothness constraint). These two distributions may be separated mainly because they differ in their regions of support in real space. From the estimated distributions the structure factor can be calculated. For deviations of particles from spherical symmetry, the excluded volume distribution may be approximated by that of an ellipsoid of revolution. Excluded volume distributions have been calculated for ellipsoids of revolution of axial ratios between 0.1 and 10 and implemented in the program *IFTc*, which is described in the appendix. The validity of the approach is demonstrated for globular particles.

1. Introduction

For several decades, small-angle scattering (SAS) has been an important tool for the investigation of structures (*e.g.* Glatter & Kratky, 1982; Feigin & Svergun, 1987). For the application of the technique to solution scattering it is frequently possible to assume that the studied system is sufficiently dilute to neglect interparticle effects, which simplifies the subsequent analysis significantly. The analysis of the measured data is usually achieved either by direct methods or by indirect methods. Using direct methods, a model is assumed for the scattering system which is then fitted to the measured data. Alternatively, by indirect Fourier transformation (IFT), the constraints from the data are combined with some additional requirements to determine a real space function which describes the scattering object. This latter approach was introduced in SAS by Glatter (1977, 1980*a*,*b*) and Glatter & Müller (1982), who used a smoothness constraint for the real space function. Other methods and constraints for IFT in SAS have been suggested (*e.g.* Moore, 1980; Svergun *et al.*, 1988; Hansen & Pedersen, 1991; Svergun, 1992; Hansen, 2000), but apart from minor modifications the most frequently used method is still that originally invented by Glatter.

For many problems the assumption of a dilute solution does not hold. In some cases the structure of interest is only to be found in non-dilute solutions. This means that experiments have to be performed with relatively concentrated samples and that the subsequent analysis of the measured data has to take interparticle effects into account.

For such non-dilute systems, the generalized indirect Fourier transformation (GIFT) extension of IFT was introduced by Brunner-Popela & Glatter (1997). In GIFT, interparticle effects are taken into account by including a structure factor in the calculations. The inclusion of the structure factor leads to a nonlinear set of equations, which must be solved either iteratively or by Monte Carlo methods.

Using GIFT the interaction between the scatterers has to be specified by the user, who has to choose a specific structure factor. On the one hand this requires some extra *a priori* information about the scattering system, but on the other hand the choice of a structure factor allows the estimation of relevant parameters describing the interaction (*e.g.* the charge of the scatterers and their interaction radius). Further input parameters may be needed, such as the temperature and the dielectric constant of the solvent. The estimation of parameters from the model may be useful (provided, of course,

that the chosen model is correct), but correlations between the parameters may also reduce the advantages of the approach (Fritz *et al.*, 2000).

In the present manuscript, an alternative method is suggested. By expressing the function describing the (real space) structure of the scatterer as a combination of an intraparticle contribution and an interparticle contribution with appropriate constraints, it is possible to separate the contributions in real space leading to the form factor and structure factor in reciprocal space.

Using this method it is not necessary to specify a structure factor to be used for the indirect transformation. Only a rough estimate of the shape of the scatterer is necessary and this estimate may be made from the scattering profile. The downside of this approach is that less detailed information may be deduced from the data. However, this is also the case for the original IFT method, which nonetheless has proved to be a most useful supplement to direct model fitting for the analysis of small-angle scattering data. Another limitation of the suggested method is that for very elongated particles it may be difficult to separate the intra- and interparticle contributions to the scattering. In this case the more general method of GIFT may be used.

Finally, a Bayesian framework provides principles for the estimation of the various hyperparameters that are necessary for IFT (*e.g.* the noise level, the maximum diameter of the scatterer and the volume fraction) and which, in their own right, may provide useful information about the scattering system.

2. Theory

2.1. Small-angle scattering

In small-angle scattering, the intensity I is measured as a function of the length of the scattering vector $q = 4\pi \sin(\theta)/\lambda$, where λ is the wavelength of the radiation and θ is half the scattering angle. For scattering from a dilute solution of monodisperse molecules of maximum dimension d, the intensity can be written in terms of the distance distribution function $p(r)$ (see Glatter, 1982):

$$I(q) = 4\pi nV \int_0^d p(r) \frac{\sin(qr)}{qr}\, \mathrm{d}r, \tag{1}$$

where n is the (average) number density of the particles and V is the volume of one particle. The distance distribution function $p(r)$ is related to the autocorrelation function $\gamma(r)$ for the scatterer as $p(r) = r^2\gamma(r)$. For uniform scattering density of the molecule, the distance distribution function is proportional to the probability distribution for the distance between two arbitrary scattering points within the molecule. For non-uniform scattering density, the distance distribution may have negative regions (if the scattering density of some regions of the scatterer is less than the scattering density of the solvent).

The simple interpretation of the distance distribution function $p(r)$ must be modified for high concentrations in order to take interparticle effects into account. The most obvious effect of an increase in concentration is usually that the calculated $p(r)$ exhibits a negative part around the maximum diameter of the scatterer. This is mainly caused by the excluded volume effect and the consequent reduction in the effective scattering length density near the scatterer. In this case, the total distance distribution function may be divided into three parts from the intra- and interparticle contributions according to (Kruglov, 2005):

$$p(r) = p_1(r) + \eta p_{\mathrm{excl}}(r) + \eta p_{\mathrm{struct}}(r), \tag{2}$$

where η is the volume fraction ($\eta = nV$), $p_1(r)$ is the distance distribution function of a single particle, $p_{\mathrm{excl}}(r)$ is the distance distribution function of the excluded volume and $p_{\mathrm{struct}}(r)$ is the remaining part of the total distance distribution function which depends on the mutual arrangement of the scatterers outside the excluded volume. For a monodisperse solution, $p_{\mathrm{excl}}(r)$ is due to the perturbation of the distribution of distances caused by the fact that the centres of two molecules cannot come closer than the minimum dimension of the molecules. At distances larger than twice the maximum dimension, $p_{\mathrm{excl}}(r) = 0$. The introduction of interparticle effects increases the integration limit of equation (1) from the maximum dimension d of a single molecule to that of the maximum length of the interaction (which may, in principle, be infinite). The first term on the right-hand side of equation (2) determines the form factor $P(q)$ when Fourier transformed according to equation (1) and the last two terms determine the structure factor $S(q)$ as specified below. Correspondingly, the intensity in equation (1) can be divided into a part that is due to intraparticle effects [the form factor $P(q)$] and a part that is due to the remaining interparticle effects [the structure factor $S(q)$],

$$I(q) \propto S(q)P(q). \tag{3}$$

For dilute solutions, $S(q) = 1$ and the measured intensity is given by the form factor $P(q)$. Equation (3) is valid for spherical monodisperse particles, but it is frequently assumed to hold true also for slightly elongated particles with some degree of polydispersity (*e.g.* Fritz & Glatter, 2006). Hayter & Penfold (1981) have given the analytical solution for $S(q)$ for a system of particles interacting through a screened Coulomb potential.

The structure factor can be written as

$$S(q) = 1 + 4\pi n \int_0^\infty h(r) r^2 \frac{\sin(qr)}{qr}\, \mathrm{d}r, \tag{4}$$

where $h(r)$ is the total correlation function (Ornstein & Zernike, 1914), which is related to the radial distribution (or pair correlation) function $g(r)$ (Zernike & Prins, 1927) for the particles by

$$h(r) = g(r) - 1. \tag{5}$$

For hard spheres, $g(r) = 0$ for $r < d$, where d is the diameter of the sphere.

Returning to equation (2), an analytical expression for the excluded volume of hard spheres was deduced by Kruglov (2005). Kruglov calculated $\gamma_{\rm excl}(r)$ and the result is shown in Fig. 1(*a*) for spheres of diameter d = 100 Å, together with $\gamma_1(r)$ and $\gamma_{\rm struct}(r)$. The corresponding distance distribution functions and the total distance distribution function $p(r)$ calculated from equation (2) are shown in Fig. 1(*b*) for the volume fraction η = 0.1. From this figure, the basis of the negative region in the total distance distribution function can be seen. The shape of $p_{\rm struct}(r)$ is shown in more detail in Fig. 1(*c*).

From Fig. 1(*b*) it is also seen that, for spheres, $p_1(r)$ and $p_{\rm struct}(r)$ have their support mainly in different regions of space. This means that, if $p_{\rm excl}(r)$ is given, it may be possible to estimate $p_1(r)$ and $p_{\rm struct}(r)$ separately from the experimental data. It is the main purpose of the present manuscript to test this assumption. As the form of $p_{\rm excl}(r)$ is only dependent upon the geometry of the particle, it requires less information for an IFT than a complete determination of the structure factor $S(q)$, which requires the specification of the interaction between the particles.

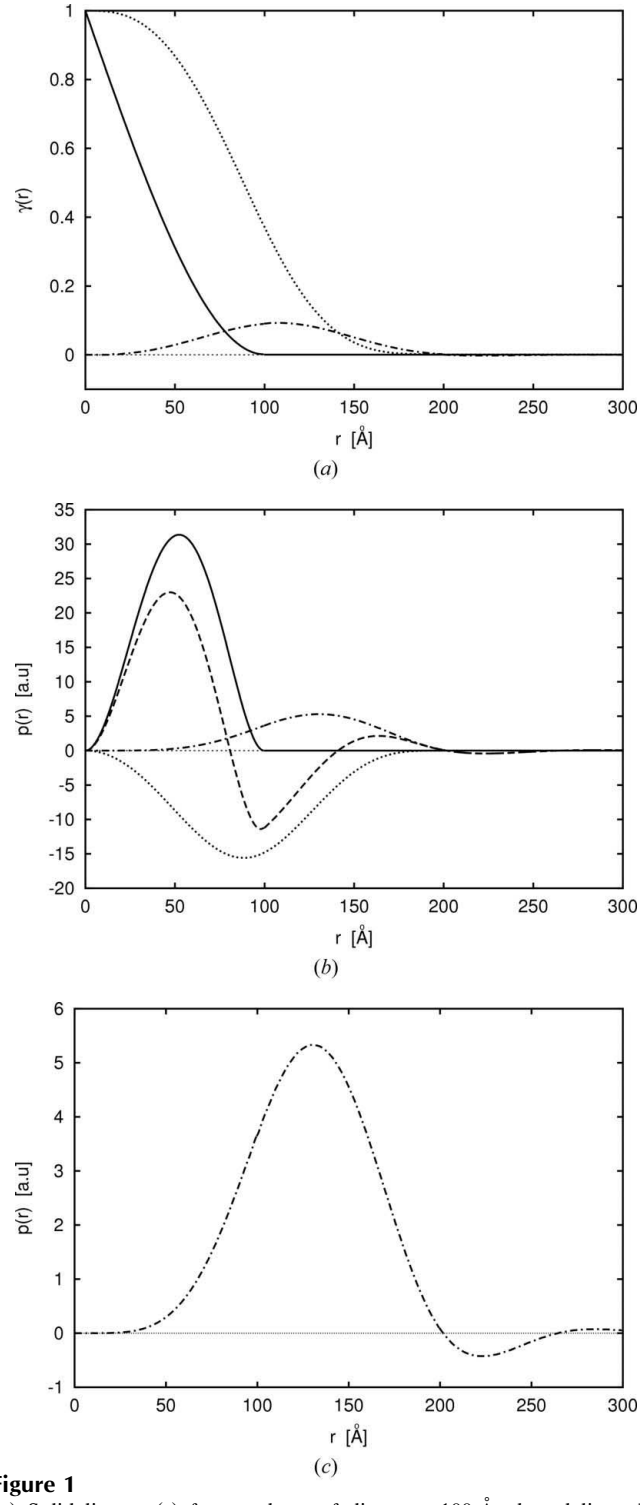


Figure 1
(*a*) Solid line: $\gamma_1(r)$ for a sphere of diameter 100 Å; dotted line: the corresponding $-\gamma_{\rm excl}(r)$; dashed–dotted line: $\gamma_{\rm struct}(r)$. (*b*) Solid line: $p_1(r)$ for a sphere of diameter 100 Å; dotted line: $p_{\rm excl}(r)$ for spheres of diameter 100 Å and volume fraction η = 0.1; dashed–dotted line: $p_{\rm struct}(r)$ for spheres of diameter 100 Å and η = 0.1; dashed line: total distance distribution function $p(r)$ according to equation (2). (*c*) $p_{\rm struct}(r)$ from (*b*), rescaled for clarity.

2.2. Estimation of the excluded volume distribution

The correlation function $\gamma_1(r) = p_1(r)/r^2$ for a homogeneous particle is proportional to the intersection volume of the particle and its ghost, shifted a distance r. To describe interactions between particles, a 'cross correlation function' $\gamma(r)_{\rm cross}$ may be introduced (Kruglov, 2005) as the intersection volume of a particle with the ghost of a second particle. Again, the ghost is shifted a distance r from the original position of the second particle. The excluded volume correlation function $\gamma_{\rm excl} = p_{\rm excl}(r)/r^2$ can then be found by integration of $\gamma(r)_{\rm cross}$ over all positions of the first particle and the ghost of the second particle, where the two 'real' particles (not the ghost) overlap. From the positions where the particles do not overlap the cross correlation function becomes $\gamma_{\rm struct} = p_{\rm struct}(r)/r^2$.

Excluded volume distributions for ellipsoids of revolution have been estimated using an approximative method similar to the method of Kruglov. Monte Carlo simulations of two ellipsoids of arbitrary orientation and separation were used to estimate the fraction of ellipsoids that overlapped as a function of their centre-to-centre separation. This was done to distinguish between the contributions to $\gamma_{\rm excl}$ and those to $\gamma_{\rm struct}$ (which is trivial for the case of spheres). For the calculation of the average intersection volume of two ellipsoids, the expressions for the correlation functions of prolate and oblate ellipsoids derived by Müller *et al.* (1996) were used. Combining these correlation functions with the outcome of the Monte Carlo simulation, an approximation to the excluded volume correlation function may be calculated following the lines of Kruglov (2005). Using $p(r) = r^2\gamma(r)$, some examples of the corresponding excluded volume distance distribution functions are shown in Fig. 2 for ellipsoids of revolution of various axial ratios.

For axial ratios a between 0.1 and 10, the calculated excluded volume distributions were parameterized and implemented in the program *IFTc*, described in Appendix *A*, which allows the axial ratio of $p_{\rm excl}(r)$ to be used as a free parameter. However, owing to the approximations used for the estimation of $p_{\rm excl}$, and the similarities in the support for p_1

and $p_{\rm excl}$ for strong deviations from spherical symmetry (as seen by comparison of Figs. 1 and 2), the program will be most reliable for globular particles.

2.3. IFT in SAS

From equation (1) for dilute scattering, the distance distribution function $p(r)$ may be approximated by $\vec{p} = (p_1, ..., p_N)$ and the measured intensity at a given q_i written as

$$I(q_i) = \sum_{j=1}^{N} a_{ij} p_j + e_i, \tag{6}$$

where e_i is the noise at data point i and $a_{ij} = 4\pi\Delta r \sin(q_i r_j)/(q_i r_j)$, where $\Delta r = r_j - r_{j-1}$. The aim of the indirect Fourier transformation is to restore $\vec{p}$, which contains the full information present in the scattering profile.

The method of Glatter (1977) is an implementation of the general method of Tikhonov & Arsenin (1977), which estimates a function of interest by minimizing a new functional written as a weighted sum of χ^2 and a regularization functional K:

$$\alpha K + \chi^2. \tag{7}$$

χ^2 is defined in the conventional manner, *i.e.*

$$\chi^2 = \sum_{i=1}^{M} \frac{[I_{\rm m}(q_i) - I(q_i)]^2}{\sigma_i^2}, \tag{8}$$

where $I_{\rm m}(q_i)$ is the measured intensity and σ_i is the standard deviation of the noise at data point i.

For the choice of the regularization functional, the expression $K = \int p''(x)^2 {\rm d}x$ is frequently used, giving preference to smooth distance distribution functions $p(r)$. Assuming $p(0) = p(d) = 0$, this regularization expression takes the discrete form

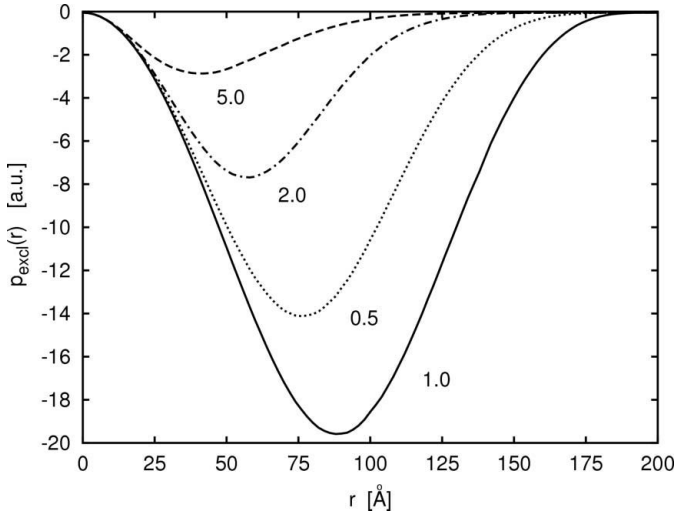


Figure 2
Solid line: $p_{\rm excl}(r)$ for a sphere of diameter 100 Å; dotted line: $p_{\rm excl}(r)$ for an oblate ellipsoid of maximum dimension 100 Å and axial ratio 0.5; dashed–dotted line: $p_{\rm excl}(r)$ for a prolate ellipsoid of maximum dimension 100 Å and axial ratio 2.0; dashed line: $p_{\rm excl}(r)$ for a prolate ellipsoid of maximum dimension 100 Å and axial ratio 5.0.

$$K = \sum_{j=2}^{N-1} \left[p_j - \frac{(p_{j-1} + p_{j+1})}{2} \right]^2 + \frac{1}{2} p_1^2 + \frac{1}{2} p_N^2, \tag{9}$$

which is similar to Glatter's original smoothness constraint.

For regularization by the maximum entropy method (Skilling, 1988), a constraint $K = \int \{p(r) \ln[p(r)/m(r)] - p(r) + m(r)\} {\rm d}r$ is used, which takes the discrete form

$$K = \sum_{j=1}^{N} p_j \ln(p_j/m_j) - p_j + m_j, \tag{10}$$

where $(m_1, ..., m_N)$ is a prior estimate of $(p_1, ..., p_N)$. Using this method will bias the estimate towards the prior, *i.e.* for the case of no constraints from the experimental data, minimizing equation (7) will lead to $\vec{p} = \vec{m}$.

As noted by Steenstrup & Hansen (1994) for $\vec{p} \simeq \vec{m}$, a second-order Taylor approximation of equation (10) will lead to

$$K \simeq \sum_{j=1}^{N} [(p_j - m_j)^2 / 2m_j]. \tag{11}$$

From this equation, it can be seen that, using the prior $m_j = (p_{j+1} + p_{j-1})/2$, the maximum entropy constraint corresponds to the smoothness constraint of equation (9) in a new metric defined by the denominator $2m_j$ in equation (11). Using this metric will combine the positivity constraint of equation (10) with the smoothness constraint of equation (9).

For non-dilute solutions, p in equation (1) is replaced by the sum of p_1, $\eta p_{\rm excl}$ and $\eta p_{\rm struct}$. For the regularization of p_1, equation (11) is used with $m_j = (p_{j+1} + p_{j-1})/2$ as mentioned above, while the conventional constraint of equation (9) is used for $p_{\rm struct}$. For the test examples shown here, only the excluded volume for a sphere is used for $p_{\rm excl}$. This implies that the diameter of the sphere from which $p_{\rm excl}$ is calculated should be very close to the maximum dimension d of p_1, but this may be generalized. This is done in the program *IFTc*, which uses $p_{\rm excl}$ for an ellipsoid of revolution, estimating the axial ratio a for the ellipsoid from the scattering data as described below.

As an additional constraint, it is assumed throughout these calculations that $p_{\rm struct} \simeq 0$ for $r < 0.5d$, in accordance with Fig. 1.

Furthermore, as $S(q) \rightarrow 1$ for $q \rightarrow \infty$, equation (3) gives

$$I(q) \propto S(q)P(q) \rightarrow P(q) \quad \text{for} \quad q \rightarrow \infty, \tag{12}$$

which can be written

$${\rm FT}[p_1(r) + p_{\rm excl}(r) + p_{\rm struct}(r)] \rightarrow {\rm FT}[p_1(r)] \quad \text{for} \quad q \rightarrow \infty, \tag{13}$$

where FT denotes the Fourier transform of equation (1). Consequently, it must hold that

$${\rm FT}[p_{\rm excl}(r) + p_{\rm struct}(r)] \rightarrow 0 \quad \text{for} \quad q \rightarrow \infty, \tag{14}$$

which may also be used in the estimation of $p_{\rm struct}$.

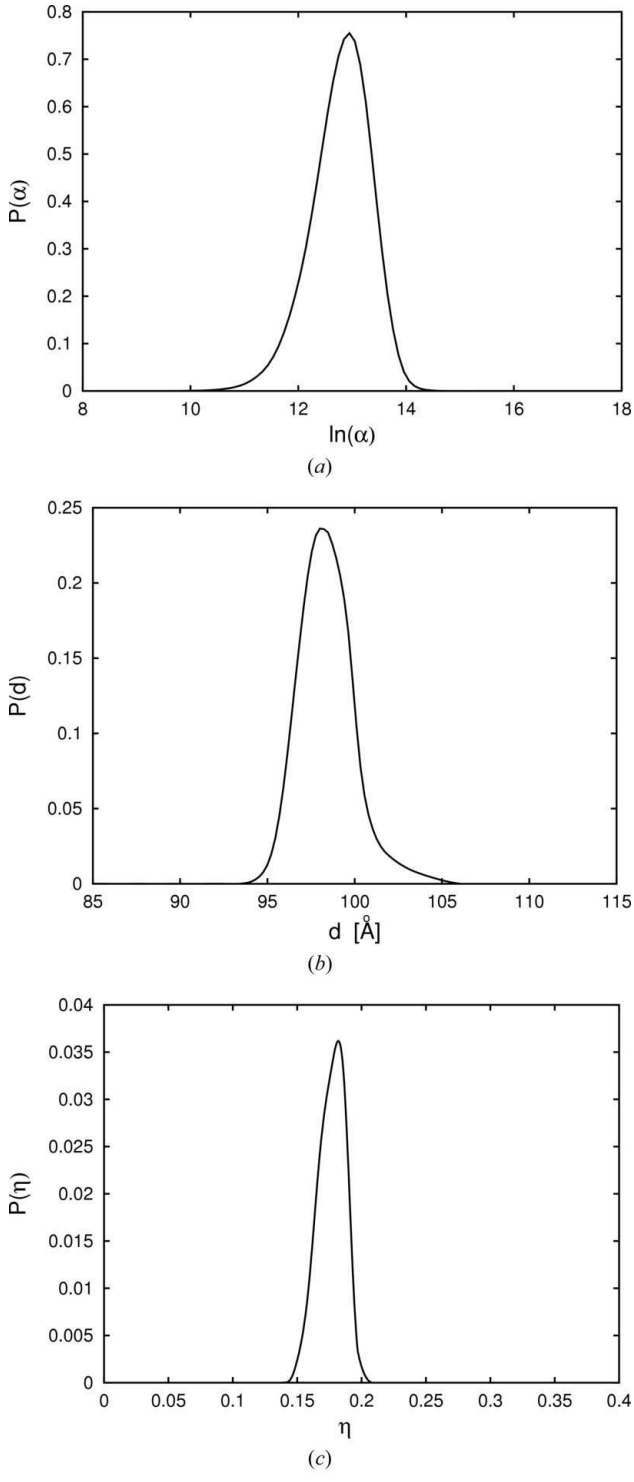


Figure 3
Posterior probabilities for hyperparameters from Fig. 4. (*a*) Probability for the Lagrange multiplier α. (*b*) Probability for the maximum diameter d. (*c*) Probability for the volume fraction η.

2.4. Bayesian analysis

The Lagrange multiplier α (which depends on the noise level of the experimental data), the maximum diameter d of the scatterer, the volume fraction η and the axial ratio a are all hyperparameters, which can be estimated from their posterior probability P for a set (α, d, η, a) after data have been measured. This probability is calculated using Gaussian approximations around the optimal estimate $\vec{p}_{\rm opt}$ for a given set of hyperparameters (α, d, η) and integrating over all solutions $\vec{p}$ for this particular set of hyperparameters (Gull, 1989; MacKay, 1992). Using the regularization from equation (9), writing $\mathbf{A} = \nabla\nabla K$ and $\mathbf{B} = \nabla\nabla\chi^2/2$, the probability of a set of hyperparameters may be written (Hansen, 2000):

$$P(\alpha, d, \eta, a) \propto \frac{\exp(-\alpha K - \chi^2/2)}{\det^{1/2}(\mathbf{A} + \alpha^{-1}\mathbf{B})}. \qquad (15)$$

In equation (15), both matrices and $(-\alpha K - \chi^2/2)$ must be evaluated at the point $\vec{p}$ where $(-\alpha K - \chi^2/2)$ takes its maximum value.

Using equation (15), the most likely value for each hyperparameter can be found from the optimum, and an error estimate for the hyperparameters can be provided from the width of the probability distribution (see Fig. 3). As the Bayesian framework ascribes a probability to each calculated solution $\vec{p}$, an error estimate for the (average) distribution of interest is provided from the individual probabilities of all the solutions (each solution corresponding to a specific choice of noise level, volume fraction and maximum dimension of the scatterer).

For a recent overview paper on Bayesian inference in physics, see Dose (2003).

3. Results

For evaluation of the suggested method, various parameters have been calculated in the examples given below, where the axial ratio a of the excluded volume distribution employed has been fixed at 1.0.

The *Guinier radius* $R_{\rm g}$ was calculated from the estimate of p_1 according to the formula (*e.g.* Glatter, 1982)

$$R_{\rm g}^2 = \frac{\int p_1(r) r^2 \, {\rm d}r}{2\int p_1(r) \, {\rm d}r}. \qquad (16)$$

The *volume fraction* η was estimated from

$$\eta \simeq \frac{-\int p_{\rm excl}(r) \, {\rm d}r}{8\int p_1(r) \, {\rm d}r}, \qquad (17)$$

deduced from the corresponding ratio for homogeneous spheres (where the excluded volume is $2^3 = 8$ times that of the sphere). Finally, the *forward scattering* $I(0)$ was determined from

$$I(0) = 4\pi n V \int_0^d p_1(r) \frac{\sin(qr)}{qr} \, {\rm d}r. \qquad (18)$$

3.1. Simulated data

For the simulated examples, M = 97 data points in the interval $q \in [0.01; 0.25]$ Å^{-1} were used. Gaussian noise was

added according to $\sigma = RI(q) + 0.01I(0)$, with R denoting the relative noise level (chosen as 1 or 4%).

Approximately $N = 200$ points were used for the estimation of $\vec{p}$. This was typically divided into 100 points for $\vec{p}_1$ in the interval $[0;d]$, and about 50 points each for $\vec{p}_{\rm excl}$ and $\vec{p}_{\rm struct}$ in the interval $[0;3d]$. Increasing N above this number only leads to an increase in CPU time for the estimation. A complete calculation may be done in a few minutes, but if necessary the algorithm used may be optimized further.

Data from spheres of radius 50 Å, from prolate ellipsoids of semi-axis (45,55) Å and from ellipsoids of semi-axis (40,60) Å were simulated for various noise levels R and volume fractions η, as shown in Table 1 and Figs. 4, 5 and 6. For all the simulated examples, the structure factor for hard spheres from Percus & Yevick (1958) was used. For the spheres, a structure factor $S(q)$ for a radius $R = 50$ Å was used, but for the ellipsoids, the radii used were $R = 48.1$ and $R = 45.8$ Å, respectively, corresponding to the (volume-)equivalent radius for a sphere. As the maximum dimensions d of the ellipsoids are 110 and 120 Å, respectively, these examples should test the limitations of the method when $2R \neq d$.

Deviations from a monodisperse structure factor were tested using a structure factor

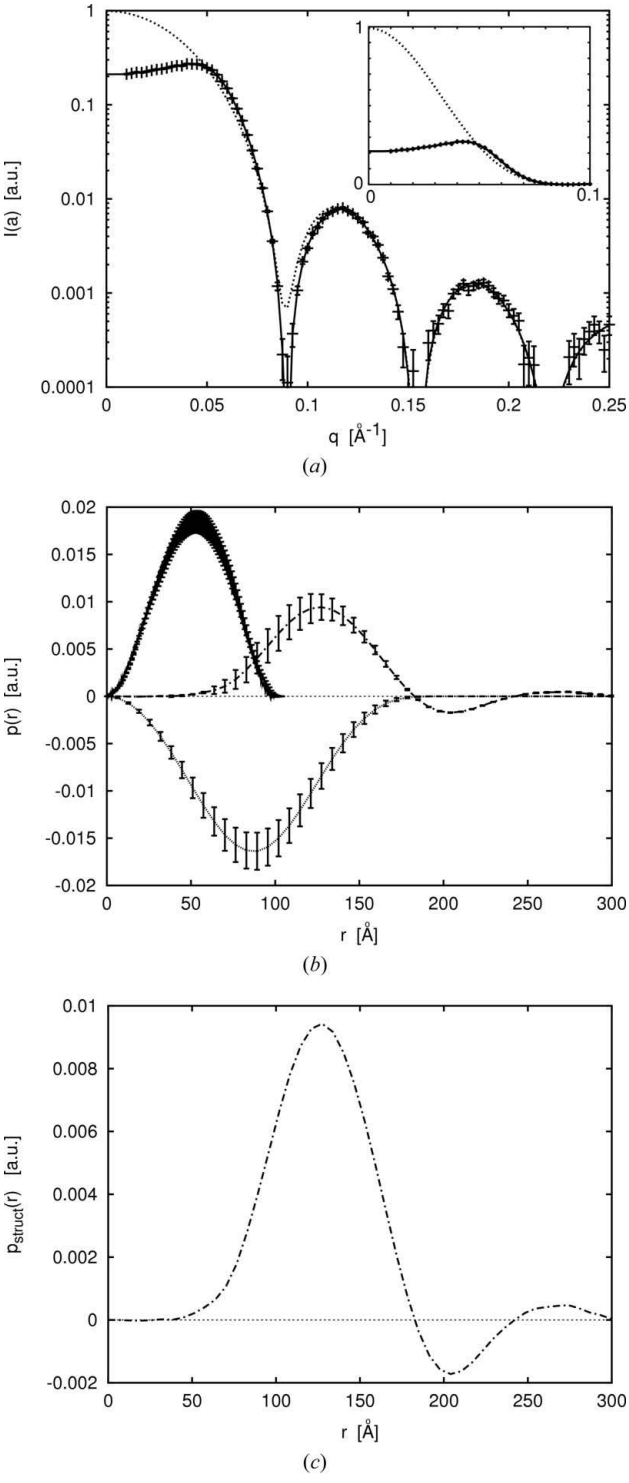


Figure 4
(*a*) Error bars: simulated data points from spheres of radius 50 Å, volume fraction 0.2 and noise 1%; solid line: fit of the data; dotted line: estimated $P(q)$. The insert shows the intensity on a linear scale for $q < 0.1$ Å^{-1}. (*b*) Error bars: $p_1(r)$. The original $p_1(r)$ is not discernable from the estimate. Dotted line: $p_{\rm excl}(r)$; dashed–dotted line: $p_{\rm struct}(r)$. (*c*) $p_{\rm struct}(r)$ rescaled.

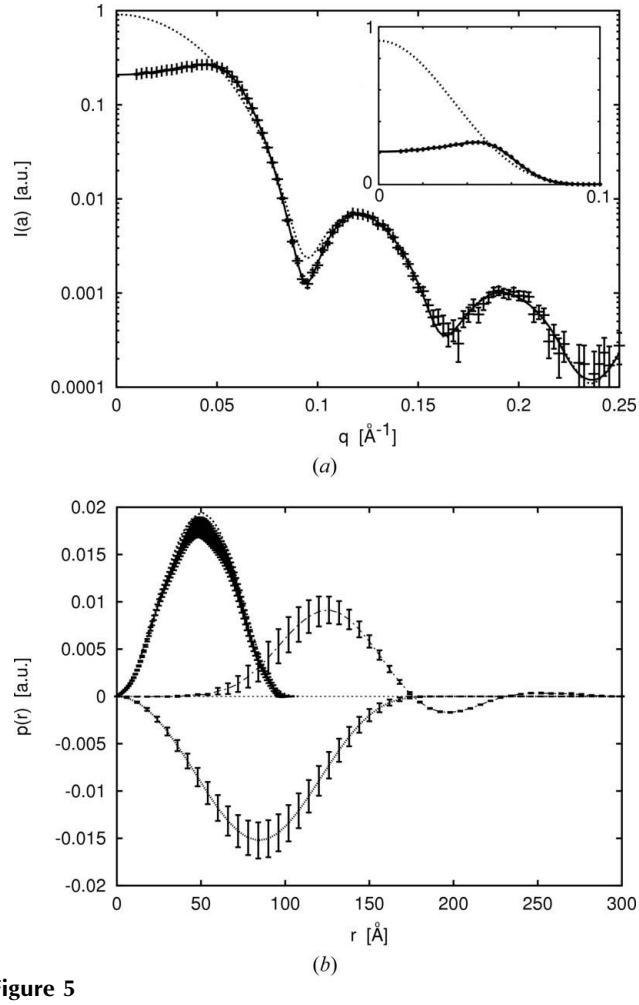


Figure 5
(*a*) Error bars: simulated data points from ellipsoids of semi-axis (45,45,55) Å, volume fraction 0.2 and noise 1%; solid line: fit of the data; dotted line: estimated $P(q)$. The insert shows the intensity on a linear scale for $q < 0.1$ Å^{-1}. (*b*) Solid line with error bars: $p_1(r)$; dotted line: original $p_1(r)$; dotted line with error bars: $p_{\rm excl}(r)$; dashed–dotted line with error bars: $p_{\rm struct}(r)$.

Table 1
Estimated parameters.

For spheres, the radius used for the simulation is shown in brackets and the polydispersity used for $S(q)$ is given in % after the brackets [according to equation (19)]. For prolate ellipsoids, the semi-axis values are shown in brackets. The estimated values in the table are given with errors, and for the simulated data the correct values for η and R_g are shown in brackets after the estimated values. The simulated examples all have $I(0) = 1$.

Example	R (%)	η (%)	R_g (Å)	d (Å)	$I(0)$ (a.u.)
Spheres (50) 0%	1	5.0 ± 0.5 (5)	38.8 ± 0.1 (38.7)	96 ± 1	1.00 ± 0.02
Spheres (50) 0%	1	18 ± 1 (20)	39.0 ± 0.3 (38.7)	98 ± 2	0.99 ± 0.06
Spheres (50) 0%	4	17 ± 2 (20)	38.5 ± 0.6 (38.7)	94 ± 3	0.95 ± 0.09
Spheres (50) 10%	4	17 ± 2 (20)	38.4 ± 0.5 (38.7)	94 ± 3	0.92 ± 0.08
Spheres (50) 20%	4	15 ± 1 (20)	38.4 ± 0.4 (38.7)	94 ± 2	0.89 ± 0.06
Ellipsoids (45,55)	1	4.7 ± 0.5 (5)	37.5 ± 0.2 (37.6)	96 ± 1	0.98 ± 0.02
Ellipsoids (45,55)	1	17 ± 1 (20)	37.4 ± 0.3 (37.6)	96 ± 2	0.96 ± 0.04
Ellipsoids (45,55)	4	17 ± 1 (20)	36.5 ± 0.5 (37.6)	91 ± 3	0.87 ± 0.06
Ellipsoids (40,60)	1	14 ± 1 (20)	33.5 ± 0.4 (36.9)	93 ± 2	0.69 ± 0.04
Ellipsoids (40,60)†	1	16 ± 1 (20)	37.6 ± 0.8 (36.9)	111 ± 2	0.99 ± 0.08
SDS 20 m*M* NaCl	–	2.8 ± 0.8	16.5 ± 0.3	50 ± 3	0.88 ± 0.03
SDS 50 m*M* NaCl	–	1.8 ± 0.9	16.5 ± 0.4	46 ± 6	0.94 ± 0.04
SDS 250 m*M* NaCl	–	1.3 ± 0.4	18.2 ± 0.3	51 ± 3	1.20 ± 0.02

† This calculation uses $d - 2R = 28$ Å.

$$S_{\rm tot}(q, \eta, R, \sigma_S) \propto \int \exp\left[\frac{(R' - R)^2}{\sigma_S^2}\right] S(q, \eta, R')\, {\rm d}R', \qquad (19)$$

where $S(q,\eta,R)$ is the structure factor for a sphere of radius R and volume fraction η, and σ_S denotes the degree of polydispersity for the total structure factor $S_{\rm tot}$. Even though equation (19) does not represent the physically correct description of polydispersity effects, it serves the purpose of testing the method for deviations from simple monodisperse structure factors. Furthermore, it has been shown that equation (19) with the structure factor of hard spheres provides an adequate approximation to the structure factors of a wide variety of non-spherical scatterers (van der Schoot, 1992; Weyerich *et al.*, 1999).

3.2. Experimental data

Small-angle neutron scattering experiments using the anionic surfactant sodium dodecyl sulfate (SDS) at a concentration of 8.65 mg ml^{-1} were carried out at the Paul Scherrer Institute in Switzerland (Arleth, 2004). The measurements were performed using Tris buffer at pH 7 and at three different ionic strengths from 20, 50 and 250 m*M* NaCl. The experimental data are shown in Figs. 7(*a*), 8(*a*) and 9(*a*), and the corresponding estimates of p_1, $p_{\rm excl}$ and $p_{\rm struct}$, using the resolution function for the specific experimental setup described by Pedersen *et al.* (1990), are shown in Figs. 7(*b*), 8(*b*) and 9(*b*). The structure factors calculated from the estimates of p_1, $p_{\rm excl}$ and $p_{\rm struct}$ are shown in Fig. 10 for all three examples.

4. Discussion

4.1. Simulated data

From the results shown in Table 1 and Figs. 4, 5 and 6, it is evident that the shape of $p_{\rm struct}$ may be reproduced from the simulated data provided that a good estimate of $p_{\rm excl}$ is used. As expected, the best results are obtained for low volume fractions and low noise levels. Polydispersities of 10 and 20% [σ_S/R according to equation (19)] were included in the calculations for the spheres shown in Table 1. The polydispersities for $S(q)$ will change the excluded volume distribution for the simulated data from that of a sphere to an average value corresponding to equation (19). The results in Table 1 indicate that small deviations from the correct excluded volume distribution do not lead to erroneous estimates. As larger discrepancies are introduced for $p_{\rm excl}$ through differences between the diameter of the spheres used for $S(q)$ and the diameter of $p_1(r)$, the deviations between the calculated and correct values for the tested parameters increase.

For ellipsoids of axial ratio 1.22, the results for the lowest volume fraction or lowest noise level are acceptable, but for ellipsoids of axial ratio 1.5 the

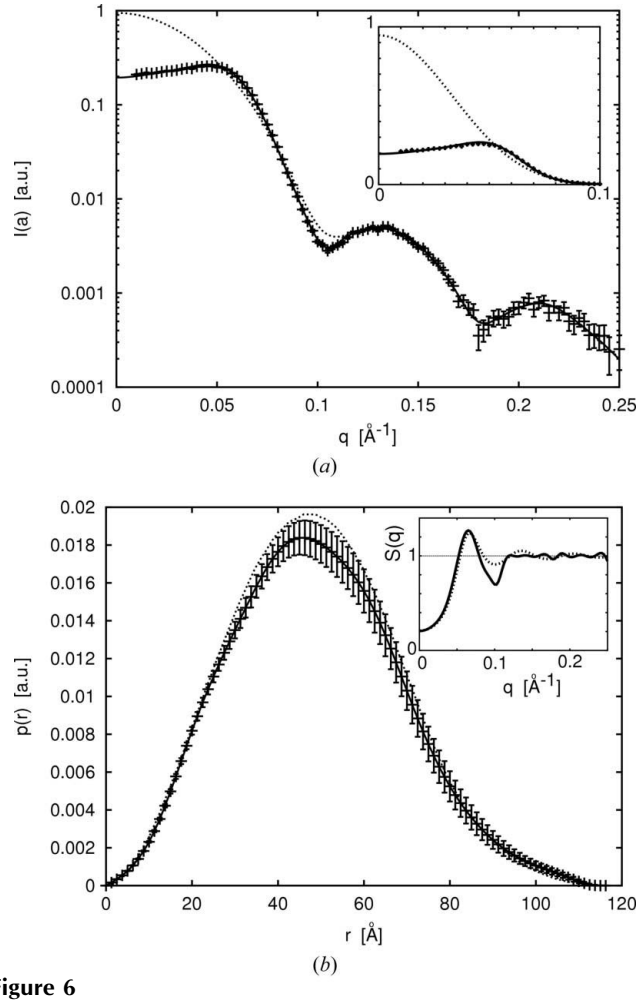


Figure 6
(*a*) Error bars: simulated data points from ellipsoids of semi-axis (40,40,60) Å, volume fraction 0.2 and noise 1%; solid line: fit of the data; dotted line: estimated $P(q)$. The insert shows the intensity on a linear scale for $q < 0.1$ Å^{-1}. (*b*) Solid line with error bars: estimated $p_1(r)$; dotted line: original $p_1(r)$. The insert shows the estimated $S(q)$ calculated as FT[$p_1(r) + p_{\rm excl}(r) + p_{\rm struct}(r)$]/FT[$p_1(r)$] and the dotted line shows the original $S(q)$.

deviations between the estimated and true values are significant. This is to be expected, as the deviation between the diameter used for $S(q)$ and the maximum dimension of p_1 is $d - 2R = 120 - 2 \times 45.8 \simeq 28$ Å or 25–30%. Entering this difference explicitly into the calculations alters the result to be in good agreement with the simulated input, as shown in Table 1 and Fig. 6.

The intra-distance distribution function p_1 usually has a smooth transition to zero around the maximum diameter of the particle. For example, for a prolate ellipsoid of revolution of semi-axis (45,55) Å, less than 1% of the total area of p_1 is found in the interval [90;110] Å and less than 0.1% in the interval [100;110] Å. This makes a reliable estimate of d very difficult in these cases, as truncation of the tail of p_1 around d will give a better value for the regularization constraint without significant implications for the Fourier transformation of p_1 and the corresponding quality of the fit of the data. To improve the estimates in such cases an additional smoothness constraint on $p_1(d)$ should be implemented.

4.2. Experimental data

The distance distribution functions p_1 in Figs. 7(b), 8(b) and 9(b) all show spherical structures with a diameter of about 5 nm. This is to be expected, as the experiments were all carried out well above the critical micelle concentration for SDS. The small tails around the maximum diameter may indicate a small amount of polydispersity in the solutions, which is also expected. The estimated volume fractions are consistent with the initial concentrations of SDS and the presence of water molecules in the micelles (more water molecules are expected to be associated with SDS at low ionic strengths). As a result of the low volume fractions, the corresponding error estimates become relatively large. At decreasing ionic strength a negative region in $p_{\rm struct}$ increases, which indicates the reduced density of micelles at this distance. The reduced density is caused by repulsion between the charged head-groups of the SDS molecules at the surface of the micelles. The structure factors calculated from the estimates of p_1, $p_{\rm excl}$ and $p_{\rm struct}$ are shown in Fig. 10.

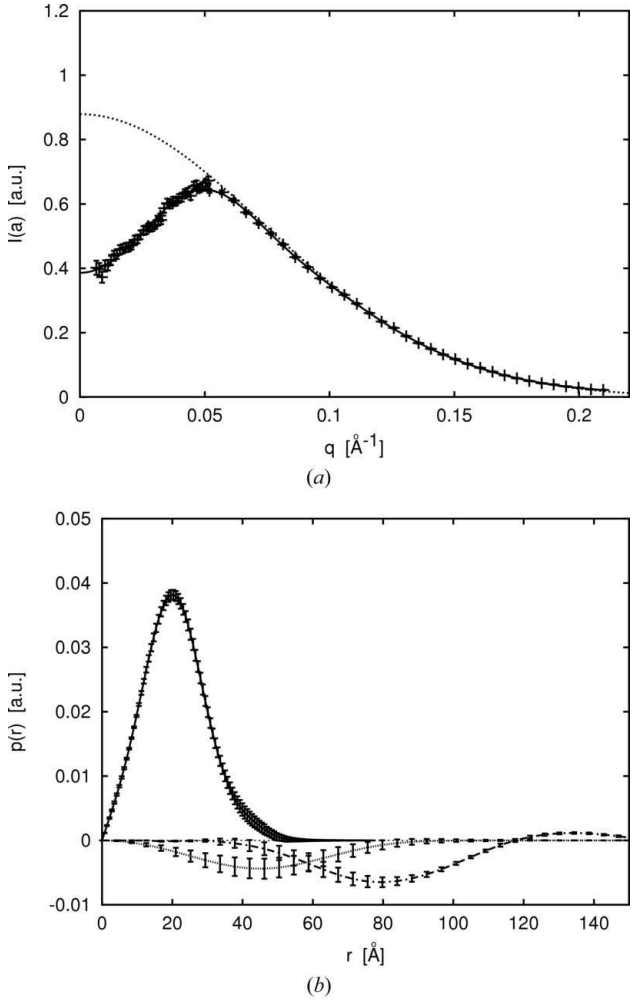


Figure 7
SDS in 20 mM NaCl. (a) Error bars: experimental data; solid line: fit of the data; dotted line: FT[$p_1(r)$]. (b) Full line with error bars: $p_1(r)$; dotted line with error bars: $p_{\rm excl}(r)$; dashed–dotted line with error bars: $p_{\rm struct}(r)$.

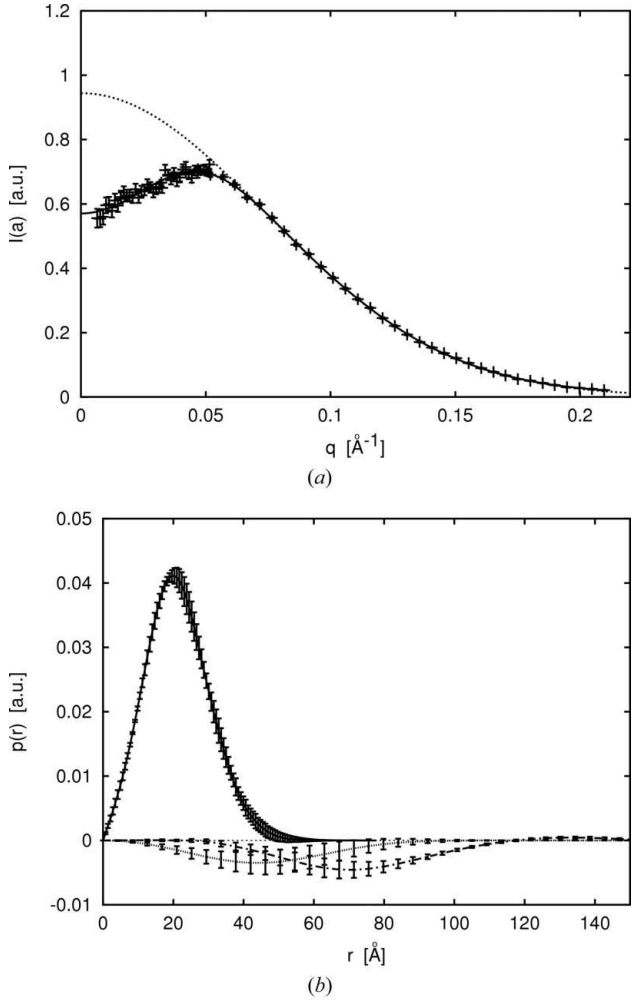


Figure 8
SDS in 50 mM NaCl. (a) Error bars: experimental data; solid line: fit of data; dotted line: FT[$p_1(r)$]. (b) Solid line with error bars: $p_1(r)$; dotted line with error bars: $p_{\rm excl}(r)$; dashed–dotted line with error bars: $p_{\rm struct}(r)$.

Comparison of Figs. 6–9 indicates an additional advantage of the free form estimation. Interpretation of data in reciprocal space is usually more difficult than the corresponding representation in real space, which is one of the reasons that $p(r)$ is usually preferred over $I(q)$. In the approach suggested here, $S(q)$ is represented by the real space distributions $p_{\rm excl}(r)$ and $p_{\rm struct}(r)$, which may allow interaction effects to be interpreted directly from the shape of $p_{\rm struct}$(r).

Analysis of the simulated and experimental data shows that it is possible to obtain simultaneous estimates of $p_{\rm struct}$ and p_1 from the data when only the shape of $p_{\rm excl}$ is assumed to be known *a priori* (the area or maximum value for $p_{\rm excl}$, as well as the overall dimension of $p_{\rm excl}$, are determined by the data).

The program used (*IFTc*) for these calculations is described in Appendix *A*. Excluded volume distributions for ellipsoids of revolution of axial ratios a between 0.1 and 10 have been implemented in the program to separate the intra- and interparticle effects for non-spherical particles. This allows a more general approach, as the axial ratio for $p_{\rm excl}$ may be estimated from the experimental data.

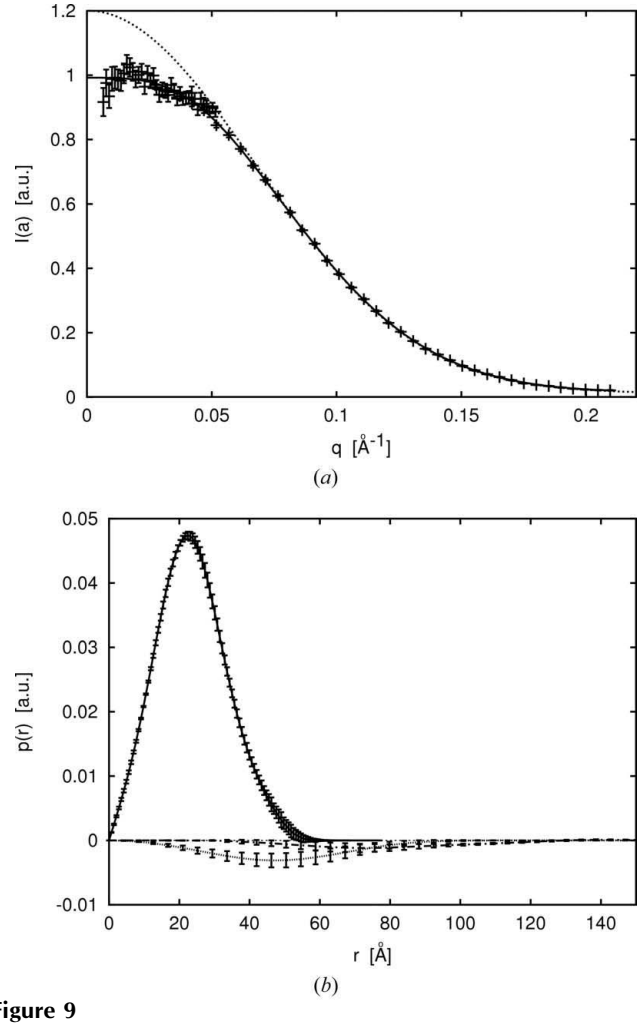


Figure 9
SDS in 250 m*M* NaCl. (a) Error bars: experimental data; solid line: fit of the data; dotted line: FT[$p_1(r)$]. (*b*) Solid line with error bars: $p_1(r)$; dotted line with error bars: $p_{\rm excl}(r)$; dashed–dotted line with error bars: $p_{\rm struct}(r)$.

Finally, it should be noted that, in order to study the limitations of the method, the volume fractions used for the simulated examples were relatively large. For lower volume fractions, the interparticle contributions are reduced and the exact form of $p_{\rm excl}$ becomes less important. Therefore, the method described here might also be used as a first-order correction for concentration effects in small-angle scattering as an alternative to the frequent practice of simply neglecting those data points that are influenced by interparticle effects.

5. Conclusion

It has been demonstrated that it is possible to separate the intra- and interparticle effects for globular particles from non-dilute solution scattering experiments using an excluded volume distribution instead of a complete structure factor for the modelling of interparticle effects. The excluded volume distribution is only dependent upon the geometry of the particles and consequently the excluded volume distribution requires less *a priori* information than the structure factor. By the introduction of the excluded volume distribution in the indirect transformation, the remaining two distributions to be estimated refer to different regions of real space, which makes their separation feasible. Therefore, interaction effects can be estimated from the data without the use of a specific model. The Bayesian estimation of the hyperparameters, including the volume fraction, allows automatic computation of the distributions of interest without any user input.

APPENDIX *A*
Description of the program *IFTc*

Algorithm. For each set of parameters (α, d, η, a) a solution (p_1, $p_{\rm excl}$, $p_{\rm struct}$) to equation (7) is calculated using the successive over-relaxation algorithm described by Steenstrup (1985). For each set of parameters the evidence given by

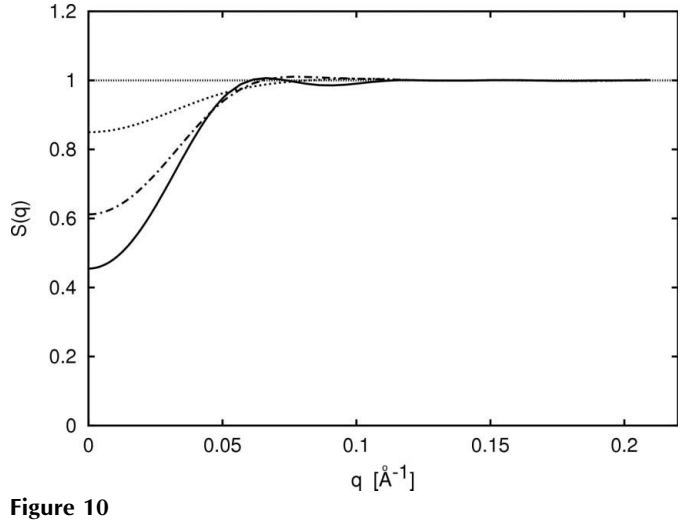


Figure 10
Calculated structure factors $S(q)$ for the SDS experiments shown in Figs. 7–9. Solid line: 20 m*M* NaCl; dashed–dotted line: 50 m*M* NaCl; dotted line: 250 m*M* NaCl.

equation (15) is calculated and the parameters are optimized (by the maximum evidence) using the Powell algorithm given by Press *et al.* (1992). When the optimum has been found, a number of solutions close to the optimum are calculated for the estimation of errors on the estimated distributions and parameters. With a 2.0 GHz processor and 1.0 Mbytes of RAM the necessary CPU time is typically 10–20 s, depending on the requirements for precision of the estimate.

Input. In addition to the data, the program requires an initial guess or starting point for the parameters (d, α, η, a). If the user does not provide this, the program makes an estimate. In principle, the program may run without any other user input than the experimental data. However, the stability of the program is improved if sensible user inputs are provided, as is the required CPU time.

Entering an initial value $\eta = 0$ will fix η at zero and the program will estimate $p(r)$ without concentration effects (as in an 'ordinary' IFT).

The value of the axial ratio a may be fixed by the user of the program. Alternatively, it may be estimated either by maximizing the evidence as a function of a, or by calculating a for an equivalent ellipsoid found using the ratio between the third and first moments of the single particle distance distribution (used as a measure of the elongation of the particle).

Owing to the low information content of traditional small-angle scattering data, as much prior information as possible should be used for the analysis, *i.e.* if the axial ratio of the molecules is known it should be entered as a constant, not a fitting parameter.

Output. The program writes estimates of $I(q)$, $P(q)$, $S(q)$ and $p_1(r)$, $p_{\rm excl}(r)$, $p_{\rm struct}(r)$ to individual files. Furthermore, the fitted parameters, including error estimates, are written to an output file. Finally, a plot file, to be used with *wgnuplot* (Williams & Kelley, 1999), is produced. The plot shows the data with the fit in linear and log plots, as well as the distance distribution functions and the structure factor.

Platform. The program has been tested under Windows XP.

Language. The source code is written in Fortran 90. The code does not need any special PC requirements.

Documentation. In addition to the program executable file, the distribution (in the form of a zip file) contains a .pdf file describing the input and output in more detail, as well as the full source code and a test example.

Availability. The distribution, including the program and the source code, can be obtained free of charge by contacting the author.

References

Arleth, L. (2004). Unpublished results.
Brunner-Popela, J. & Glatter, O. (1997). *J. Appl. Cryst.* **30**, 431–442.
Dose, V. (2003). *Rep. Prog. Phys.* **66**, 1421–1461.
Feigin, L. A. & Svergun, D. I. (1987). *Structure Analysis by Small-Angle X-ray and Neutron Scattering*. New York: Plenum.
Fritz, G., Bergmann, A. & Glatter, O. (2000). *J. Chem. Phys.* **113**, 9733–9740.
Fritz, G. & Glatter, O. (2006). *J. Phys. Condens. Matter*, **18**, S2403–S2419.
Glatter, O. (1977). *J. Appl. Cryst.* **10**, 415–421.
Glatter, O. (1980*a*). *J. Appl. Cryst.* **13**, 7–11.
Glatter, O. (1980*b*). *J. Appl. Cryst.* **13**, 577–584.
Glatter, O. (1982). *Small-Angle X-ray Scattering*, edited by O. Glatter & O. Kratky, pp. 119–165. London: Academic Press.
Glatter, O. & Kratky, O. (1982). Editors. *Small-Angle X-ray Scattering*. London: Academic Press.
Glatter, O. & Müller, K. (1982). *Makromol. Chem.* **183**, 465–479.
Gull, S. F. (1989). *Maximum Entropy and Bayesian Methods*, edited by J. Skilling, pp. 53–71. Dordrecht: Kluwer Academic Publishers.
Hansen, S. (2000). *J. Appl. Cryst.* **33**, 1415–1421.
Hansen, S. & Pedersen, J. S. (1991). *J. Appl. Cryst.* **24**, 541–548.
Hayter, J. B. & Penfold, J. (1981). *Mol. Phys.* **42**, 109–118.
Kruglov, T. (2005). *J. Appl. Cryst.* **38**, 716–720.
MacKay, D. J. C. (1992). *Maximum Entropy and Bayesian Methods, Seattle 1991*, edited by C. R. Smith, G. J. Erickson & P. O. Neudorfer, pp. 39–66. Dordrecht: Kluwer Academic Publishers.
Moore, P. B. (1980). *J. Appl. Cryst.* **13**, 168–175.
Müller, J. J., Hansen, S. & Pürschel, H.-V. (1996). *J. Appl. Cryst.* **29**, 547–554.
Ornstein, L. S. & Zernike, F. (1914). *Proc. Sect. Sci. K. Ned. Akad. Wet.* **17**, 793–806.
Pedersen, J. S., Posselt, D. & Mortensen, K. (1990). *J. Appl. Cryst.* **23**, 321–333.
Percus, J. K. & Yevick, G. J. (1958). *Phys. Rev.* **110**, 1–13.
Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (1992). *Numerical Recipes in Fortran 77. The Art of Scientific Computing*, 2nd ed., pp. 387–448. New York: Cambridge University Press.
Schoot, P. van der (1992). *Macromolecules*, **25**, 2923–2927.
Skilling, J. (1988). *Maximum Entropy and Bayesian Methods in Science and Engineering*, Vol. 1, edited by G. J. Erickson & C. R. Smith, pp. 173–187. Dordrecht: Kluwer Academic Publishers.
Steenstrup, S. (1985). *Aust. J. Phys.* **38**, 319–327.
Steenstrup, S. & Hansen, S. (1994). *J. Appl. Cryst.* **27**, 574–580.
Svergun, D. I. (1992). *J. Appl. Cryst.* **25**, 495–503.
Svergun, D. I., Semenyuk, A. V. & Feigin, L. A. (1988). *Acta Cryst.* A**44**, 244–250.
Tikhonov, A. N. & Arsenin, V. Ya. (1977). *Solution of Ill-Posed Problems*. New York: Wiley.
Weyerich, B., Brunner-Popela, J. & Glatter, O. (1999). *J. Appl. Cryst.* **32**, 197–209.
Williams, T. & Kelley, C. (1999). *GNUPLOT*. Version 3.7. http://www.gnuplot.info.
Zernike, F. & Prins, J. A. (1927). *Z. Phys.* **41**, 184–194.