Open access • Journal Article • DOI:10.1107/S0021889803000268

Uniqueness of ab initio shape determination in small-angle scattering

Vladimir Volkov, Dmitri I. Svergun

Institutions: Russian Academy of Sciences

Published on: 01 Jun 2003 - Journal of Applied Crystallography (International Union of Crystallography (IUCr))

Topics: Ab initio, Scattering and Small-angle scattering

Uniqueness of *ab initio* shape determination in small-angle scattering

Vladimir V. Volkov[a*] and Dmitri I. Svergun[b,a]

[a]*Institute of Crystallography, Russian Academy of Sciences, Moscow, Russia, and* [b]*European Molecular Biology Laboratory, c/o DESY, Hamburg, Germany. E-mail: vvo@ns.crys.ras.ru*

Scattering patterns from geometrical bodies with different shapes and anisometry (solid and hollow spheres, cylinders, prisms) are computed and the shapes are reconstructed *ab initio* using envelope function and bead modelling methods. A procedure is described to analyze multiple solutions provided by bead modeling methods and to estimate stability and reliability of the shape reconstruction. It is demonstrated that flat shapes are more difficult to restore than elongated ones and types of shapes are indicated, which require additional information for reliable shape reconsrtuction from the scattering data.

Keywords: Small-angle scattering; shape determination; uniqueness; structure modeling

1. Introduction

Small-angle scattering (SAS) patterns from monodisperse systems of non-interacting particles (e.g. from dilute protein solutions) are isotropic functions $I(s)$ of the momentum transfer $s = 4\pi\sin\theta/\lambda$, where 2θ is the scatting angle and λ is the radiation wavelength. Due to chaotic positions of individual particles, $I(s)$ is proportional to the scattering from a single particle averaged over all orientations (Feigin & Svergun, 1987). *Ab initio* analysis aiming at recovering three-dimensional structure from one-dimensional scattering curve is obviously ambiguous, as many different models may yield the same SAS curve with near the same accuracy. Homogeneous approximation is often used to constrain the solution by assuming uniform density inside the particle and discarding the influence of its internal structure. Such a simplification can be employed e.g. in the analysis of low resolution (to about s=2-3 nm^{-1}) portions of X-ray scattering patterns from sufficiently large (>50 KDa) proteins.

In the past, trial-and-error method was employed for shape modeling by computing the scattering patterns from different shapes to compare them with the experimental data. One could distinguish between two modeling strategies. One strategy was to keep the number of model parameters as low as possible, using three-parameter bodies like prisms, ellipsoids or cylinders; the other was to construct complicated bodies from collections of spheres (*i.e.* to use many parameters) and to restrain the model by additional information (Kratky & Pilz, 1978). Evolution of the two strategies and improved power of computers lead to the modern *ab initio* shape determination methods. These methods are now widely employed in practice by different research groups, e.g. (Bada *et al.*, 2000; Grossmann *et al.*, 2000; Bernocco *et al.*, 2001; Egea *et al.*, 2001; Scott *et al.*, 2002) but the question of uniqueness of the models is rarely discussed. In the present paper, the problem of uniqueness of *ab initio* shape analysis is addressed by model calculations on particles of different shape and anisometry, and practical recommendations are given for assessing the reliability of the shape reconstruction.

2. Shape determination methods

In the first general *ab initio* method based on a few parameters approach, Stuhrmann (1970) proposed to represent the particle shape by an angular envelope function $r=F(\omega)$, where (r,ω) are spherical coordinates. The particle density was unity inside the envelope and zero outside. The envelope was described by a series of spherical harmonics:

$$F(\omega) = \sum_{l=0}^{L} \sum_{m=-l}^{l} f_{lm} Y_{lm}(\omega) \qquad (1)$$

where the maximum order of harmonic L defined the resolution. The low resolution shape is thus defined in a general case by a $(L+1)^2$ -6 parameters, which can determined by a non-linear minimization procedure to fit the scattering data. This approach was further developed by (Svergun & Stuhrmann, 1991) who proposed algorithms for rapid computation of scattering intensities from such a model and implemented in the computer program SASHA (Svergun *et al.*, 1996). It was demonstrated that in practice a unique envelope can be extracted from the scattering data up to the resolution L=4.

The modeling using angular envelope function has limitations in describing complicated, e.g. very anisometric, particles, or those having internal cavities. Although the solution is unique (of course, up to an enantiomorphic shape, which always provides the same intensity), series (1) may not ensure adequate representation of the shape leading to systematic errors (even if the scattering data is neatly fitted). A more comprehensive description is achieved in the bead modeling methods, which use the improved speed of modern computers to revive the strategy of many parameter modeling in different flavors of Monte Carlo-type search. The *ab initio* bead modeling in a confined volume was first proposed by (Chacon *et al.*, 1998; Chacon *et al.*, 2000). The maximum dimension $D_{\max}$ of a particle is readily obtained from the scattering pattern and the particle must obviously fit inside a sphere of this diameter. If one fills the sphere with M densely packed beads (spheres of radius $r_0 << D_{\max}$), each of these beads may belong either to the particle (X_i =1) or to the solvent (X_i =0), and the particle shape is described by a string, $\boldsymbol{X}$, of M bits. Scattering intensity from the bead model is computed e.g. using Debye's formula (Debye, 1915)

$$I(s) = \mathrm{f}^2(s) \sum_{i=1}^{M} \sum_{j=1}^{M} X_i X_j \frac{\sin(sr_{ij})}{sr_{ij}} \qquad (2)$$

where $r_{ij} = |\boldsymbol{r}_i - \boldsymbol{r}_j|$ is the distance between the beads and $f(s)$ is the bead form factor (scattering amplitude from a sphere of radius r_0). Starting from a random distribution of 1 and 0, the model is modified to find the binary string (i.e. the shape) that fits the experimental data using a genetic algorithm. In a more general approach (Svergun, 1999), the beads may belong to different components so that the shape and internal structure of multicomponent particles (e.g. nucleoproteins) can be reconstructed by simultaneously fitting scattering data at different contrasts (Svergun & Nierhaus, 2000). For single component particles, the procedure degenerates to an *ab initio* shape determination. The model intensity is computed using spherical harmonics to speed up the computations and compactness and connectivity constraints are imposed in the search, implemented in the simulated annealing program DAMMIN (Svergun, 1999). *Ab initio* Monte Carlo-type approaches without limitation of the search space are also available (program SAX3D, (Walther, Cohen & Doniach, 2000) and the program SASMODEL (Vigil *et al.*, 2001)).

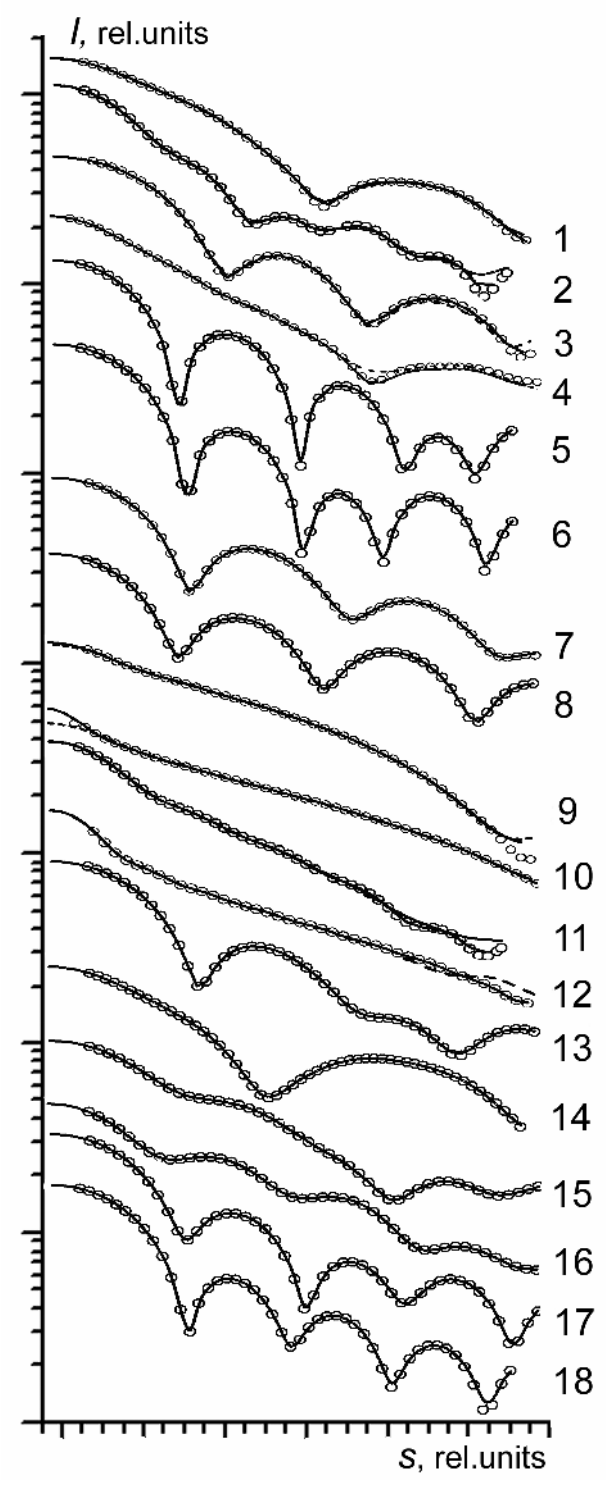

Figure 1

Scattering curves computed from the model bodies (circles) and fits by SASHA (dashed lines) and DAMMIN (solid lines). The numbers correspond to those in Figs. 2-5. The abscissae and ordinates of the individual curves are multiplied by appropriate scale factors for better visualization.

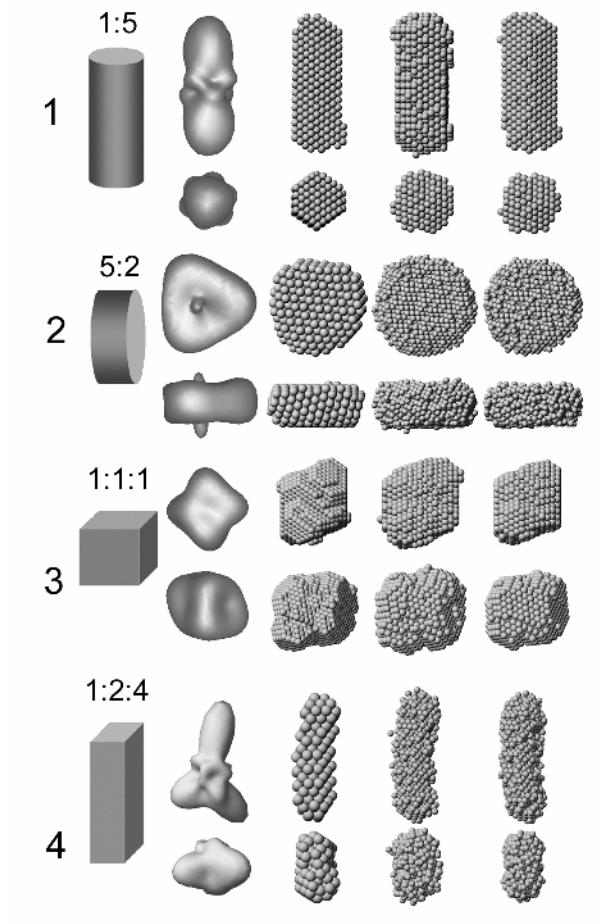

Figure 2

Shape determination of globular solid particles. Here and below, the models (from left to right) are: the geometrical body to be restored; SASHA envelope model (if applicable); one or two typical DAMMIN models; the two rightmost panels always display the results of DAMAVER: TSR (left) and MPV (right). The models (except those with spherical shape) are displayed in two orthogonal views. All bodies are marked by successive numbers (left column) through Fig. 2 to 5. Structure proportions are indicated at the top of each geometrical body as ratio of diameter to hight (solid cylinders), ratio of edges (prisms), ratio of inner and outer diameters (hollow spheres), ratio inner diameter - outer diameter - hight (hollow cylinders).

The search models employed in all Monte-Carlo based approaches are described by hundreds or thousands parameters (e.g. occupancy indices in bead modelling). Running these programs several times on the same data starting from different initial approximations may yield different final models, and the question arises, how reliable is the many-parameter approach. The authors of the shape determination programs usually present their successful applications to simulated and practical examples. The present paper is an attempt to explore limits of *ab initio* shape determination by performing model calculations on geometrical bodies to find out which shapes and structural details can and which cannot be reliably determined from the scattering data.

3. Model calculations

The geometrical bodies taken for the simulations differed by shape (spheres, cylinders and prisms) and anisometry. Elongated and flattened particles (ratio of length to width 1:C and C:1, respectively, $1<C<10$) and hollow particles with the ratio of the inner radius to the outer radius $0.33 < r/R < 0.66$ were considered. Scattering patterns

were calculated using analytical equations (Feigin & Svergun, 1987) in the angular range covering about 15 to 18 Shannon channels (the width of a Shannon cnannel is $\Delta s=\pi/D_{max}$ (Shannon & Weaver, 1949)). This ensured approximately the same information content in the simulated data for different shapes (Moore, 1980). The computed curves were used to reconstruct the shape *ab initio* using the simulated annealing program DAMMIN (Svergun, 1999), and, for the shapes without cavities, also using the envelope program SASHA (Svergun et al., 1996). The results provided by the *ab initio* programs were compared to the actual shapes. Both programs were run in batch modes using default answers without symmetry restrictions. For SASHA, multipole resolutiomn of L=6 was used; all DAMMIN calculations were made inside the search volume with diameter $1.05*D_{max}$; 'fast' mode was used for globular particles and 'slow' mode for anisometric particles.

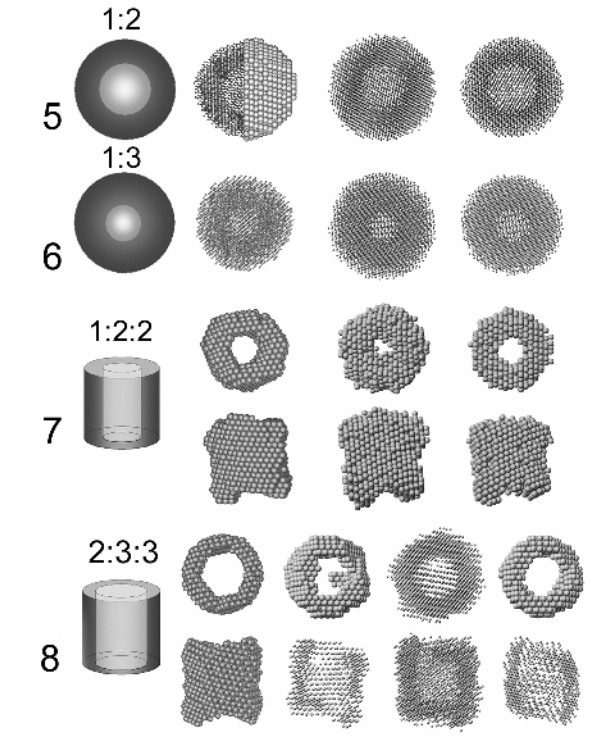

Figure 3

Shape determination of hollow globular particles. See annotation to Fig. 2 for details.

The envelope determination technique (program SASHA) gives a single solution. In contrast, DAMMIN provides many solutions (spatial distributions of beads) for runs with different seeds for random number generator (i.e. with randomly generated starting models). Analysis of the DAMMIN solutions yielding nearly identical scattering patterns can serve as an indicator of the stability of the solution. For automated analysis of independent DAMMIN reconstructions, a program package DAMAVER was written based on the program SUPCOMB (Kozin & Svergun, 2001). The latter program aligns two arbitrary low or high resolution models represented by ensembles of points by minimizing a dissimilarity measure called normalized spatial discrepancy (NSD). For every point (bead or atom) in the first model, the minimum value among the distances between this point and all points in the second model is found, and the same is done for the points in the second model. These distances are added and normalized against the average distances between the neighboring points for the two models. Generally, NSD values close to one indicate that the two models are similar.

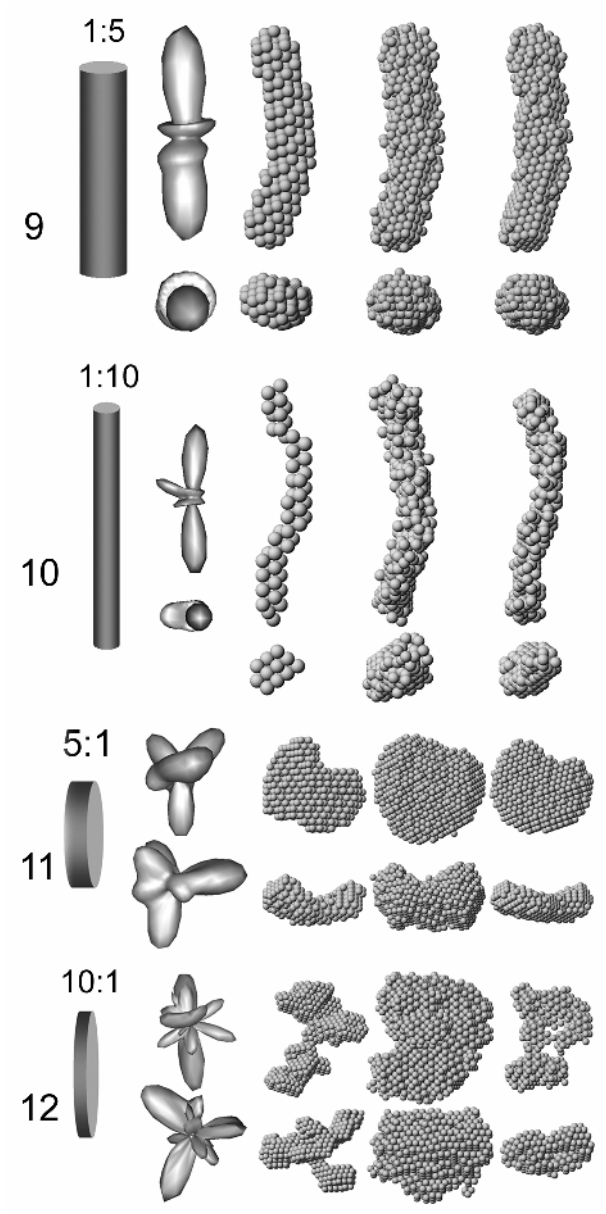

Figure 4

Shape determination of anisometric solid particles. See annotation to Fig. 2 for details.

For each model body, ten independent DAMMIN reconstructions were analyzed by DAMAVER as follows. The values of NSD were computed between each pair in the set and a mean value over all pairs <NSD> and dispersion Δ(NSD) were calculated. For each reconstruction, the average value NSD_k with respect to the rest of the set was computed and the reference reconstruction with lowest NSD_k was selected. Possible outliers with NSD_k exceeding <NSD> + 2 Δ(NSD) were discarded. All the models except the outliers were superimposed onto the reference model using SUPCOMB and the entire assembly of beads was remapped onto a densely packed grid of beads where each grid point was characterized by its occupancy factor (the number of beads in the entire assembly that are in the vicinity of the grid point). The grid points with non-zero occupancy form a total spread region (TSR),

and a portion of the TSR with higher occupancies was selected (most populated volume, MPV) to yield the volume equal to the average excluded volume of all the reconstructions. The scattering computed from the MPV would not fit the experimental data but this model should preserve the most probable features of the solution.

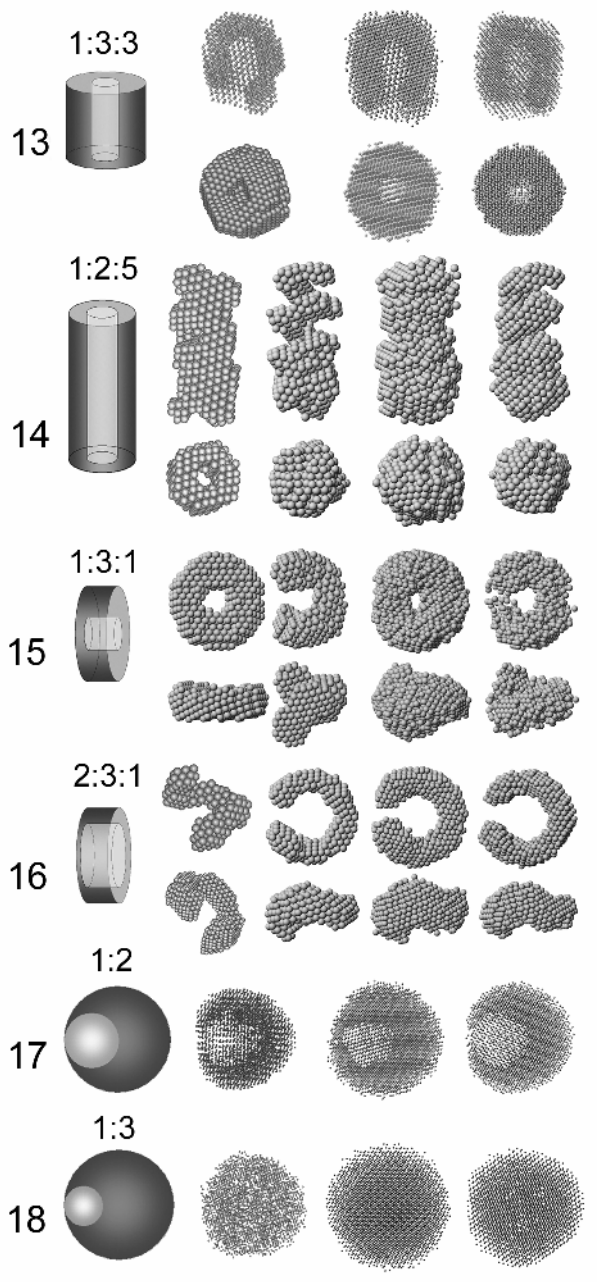

Figure 5

Shape tetermination of hollow anisometric and acentric particles. See annotation to Fig. 2 for details.

4. Results

The scattering patterns from the model bodies are presented in Fig. 1 along with the fits provided by the *ab initio* reconstructions (which were in most cases undistinguishable from the theoretical scattering patterns).

Fig. 2 demonstrates that solid bodies with moderate anisometry (elongated particles up to 1:5 and flattened up to 5:2) can be reliably reconstructed from the scattering data. The shapes obtained by SASHA reasonably represent the overall anisometry, albeit displaying artificial features due to limited resolution of series (1). DAMMIN yields very stable reconstructions, which is reflected in the mean value <NSD> of about 0.4-0.7 for all these cases. Hollow globular models can also be well reconstructed (Fig. 3). For a hollow concentric sphere, even rather small voids with r/R = 0.33, are clearly revealed, and the shape of a hollow cylinder with r/R = 0.5 is also neatly restored.

Shape reconstructions of anisometric particles are less stable and reliable. For elongated bodies, anisometry 1:5 is limiting for SASHA whereas DAMMIN still represents an elongated particle with the ratio 1:10, but tends to provide a slightly bent shape, even after the averaging procedure (Fig. 4, bodies 9 and 10 , <NSD> = 0.5 and 0.6, respectively). Flattened particles represent yet more difficult case, and starting from the anisometry 5:1, the shapes provided by SASHA are meaningless. The individual solutions from DAMMIN also show artifacts but for the anisometry 5:1 (Fig. 4, body 11), <NSD> = 0.75) the MPV reasonably well represents the flat initial shape. For the anisometry 10:1, the TSR is very large (<NSD> = 1.3) and even the MPV does not resemble a disk-shaped particle (Fig. 4, body 12).

Elongated hollow particles with higher anisometry or narrow channels (Fig. 5, bodies 13,14) may also pose problems for the shape reconstruction. Although the DAMMIN solutions are stable (<NSD> about 0.45-0.65), the channels may appear closed from one or both sides, in individual solutions and also in the MPV. Similar to what was observed for solid models, hollow flattened particles are even more difficult to restore: for different r/R ratios, the resulting shapes may show a helical turn instead of a hollow disk, even after the averaging (Fig. 5, bodies 15,16). Acentric voids in hollow spheres are only reconstructed if r/R is about 0.5 (Fig. 5, body 17); smaller voids are just "dispersed" inside the entire spherical volume (Fig. 5, body 18). For globular particles with small cavities, the <NSD> value may be small, (typically 0.4-0.6) indicating overall similarity of the models, but the details of the internal structure may significantly differ between the solutions.

5. Discussion

In the present study, we tried to find out simple geometrical shapes, which cannot be reliably restored from the small-angle scattering data. That is why we used relatively wide ranges of the scattering vectors and did not add noise to the simulated profiles. If a shape cannot be reliably restored under ideal conditions, it is unrealistic to hope and restore it *ab initio* from real experimental data without additional information. The bodies with sharp edges (cylinders and prisms) were deliberately taken instead of ellipsoids, because the edges are generally more difficult to reconstruct.

We have performed extensive computations on geometrical bodies with different shapes, anisometries and internal cavities; this paper presents a selection of most representative results. The main conclusions from the model calculations are:

(i) Flattened particles are generally more difficult to reconstruct than the elongated ones. The degree of anisometry for a reliable *ab initio* shape reconstruction should not exceed 1:10 and 5:1 for elongated and flattened particles, respectively. Not surprisingly, it was found that for anisometric particles the bead modeling is superior over the envelope determination technique. In general, it is advisable to check the particle anisometry, e.g. by finding a best-fit three-axial ellipsoid - the option provided, in particular, in the program SASHA - prior to running a shape determination program.

(ii) For globular and elongated particles, internal cavities as small as 1/3 of diameter or cross-section can be restored. For flat hollow particles, the general shape is reconstructed but artifacts may appear (e.g a disk may evolve to a helical turn). For spherical particles, concentric voids are better reconstructed than acentric ones.

(iii) The mean NSD between independent reconstructions obtained by bead modeling provide a useful estimate of the reliability of the solution. The values of <NSD> exceeding 0.7 yield large TSR and indicate that the reconstruction is unstable; in these cases additional information is required for reliable shape determination.

(iv) Averaging of the independent reconstructions allows one to enhance the most persistent features of the bead models and in most cases improves the quality of the shape reconstruction. There could, however, be cases, when the averaging has little effect and the MPV still shows systematic deviations from the initial shape (Fig. 2, body 2; Fig. 5, bodies 15,16).

The programs SASHA and DAMMIN were used for the simulations because they were written by the authors. Calculations with the programs DALAI_GA (Chacon et al., 1998; Chacon et al., 2000) and SAXS3D (Walther, Cohen & Doniach, 2000) yielded the results similar to DAMMIN (somewhat less stable, but this might be attributed to not necessarily optimal choice of parameters). All computations in the present paper were made without imposing symmetry restrictions on possible models, and without information about particle anisometry (in particular, for DAMMIN, within spherical search volumes. Although both SASHA and DAMMIN allow one to impose restrictions both on symmetry and anisometry, these restrictions were not used to keep the results as general as possible. It must thus be stressed that our conclusions refer to "pure" *ab initio* shape determination without any additional information. The use of symmetry allows one to reliably restore even highly flattened particles, (see e.g. model calculations by Volkov et. al., this issue).

The programs SASHA and DAMMIN, along with the averaging package DAMAVER, are available from www.embl-hamburg.de/ExternalInfo/Research/Sax. The program DAMAVER can be used to analyse the stability and construct average models provided by different *ab initio* methods, in particular also by the dummy residues method of (Svergun, Petoukhov & Koch, 2001)

The authors acknowledge the financial support provided by the Russian Foundation for Basic Research, grant No 01-02-17040, and by the INTAS grant No 00-243.

References

Bada, M., Walther, D., Arcangioli, B., Doniach, S. & Delarue, M. (2000). *J Mol Biol* **300**, 563-574.
Bernocco, S., Finet, S., Ebel, C., Eichenberger, D., Mazzorana, M., Farjanel, J. & Hulmes, D. J. (2001). *J Biol Chem* **276**, 48930-48936.
Chacon, P., Diaz, J. F., Moran, F. & Andreu, J. M. (2000). *J Mol Biol* **299**, 1289-1302.
Chacon, P., Moran, F., Diaz, J. F., Pantos, E. & Andreu, J. M. (1998). *Biophys J* **74**, 2760-2775.
Debye, P. (1915). *Ann. Physik* **46**, 809-823.
Egea, P. F., Rochel, N., Birck, C., Vachette, P., Timmins, P. A. & Moras, D. (2001). *J. Mol. Biol.* **307**, 557-576.
Feigin, L. A. & Svergun, D. I. (1987). *Structure analysis by small-angle x-ray and neutron scattering*. New York, Plenum Press.
Grossmann, J. G., Hasnain, S. S., Yousafzai, F. K. & Eady, R. R. (2000). *J Biol Chem*.
Kozin, M. B. & Svergun, D. I. (2001). *J. Appl. Crystallogr.* **34**, 33-41.
Kratky, O. & Pilz, I. (1978). *Quarterly Reviews of Biophysics* **11**, 39-70.
Moore, P. B. (1980). *J. Appl. Cryst.* **13**, 168-175.
Scott, D. J., Grossmann, J. G., Tame, J. R., Byron, O., Wilson, K. S. & Otto, B. R. (2002). *J Mol Biol* **315**, 1179-1187.
Shannon, C. E. & Weaver, W. (1949). *The mathematical theory of communication.* Urbana, University of Illinois Press.
Svergun, D. I. (1999). *Biophys J* **76**, 2879-2886.
Svergun, D. I. & Nierhaus, K. H. (2000). *J Biol Chem* **275**, 14432-14439.
Svergun, D. I., Petoukhov, M. V. & Koch, M. H. J. (2001). *Biophys J* **80**, 2946-2953.
Svergun, D. I. & Stuhrmann, H. B. (1991). *Acta Crystallogr.* **A47**, 736-744.
Svergun, D. I., Volkov, V. V., Kozin, M. B. & Stuhrmann, H. B. (1996). *Acta Crystallogr.* **A52**, 419-426.
Vigil, D., Gallagher, S. C., Trewhella, J. & Garcia, A. E. (2001). *Biophys J* **80**, 2082-2092.
Walther, D., Cohen, F. E. & Doniach, S. (2000). *J. Appl. Crystallogr.* **33**, 350-363.