conference papers

Small-angle scattering studies of macromolecular solutions

Dmitri I. Svergun

European Molecular Biology Laboratory, Hamburg Outstation, EMBL c/o DESY, Notkestrasse 85, D-22603 Hamburg, Germany, and Institute of Crystallography, Russian Academy of Sciences, Leninsky pr. 59, 117333 Moscow, Russia. Correspondence e-mail: svergun@embl-hamburg.de

In recent years, major progress has been achieved in developing novel approaches to interpret small-angle scattering data from solutions of biological macromolecules in terms of three-dimensional models. These advanced methods include: *ab initio* low-resolution shape and domain structure determination; modelling of quaternary structure by rigid-body refinement; simultaneous analysis of multiple scattering patterns, *e.g.* from contrast variation in neutron scattering to study multicomponent complexes; validation of high-resolution models; and addition of missing loops and domains. The new techniques will be presented and practical applications of the methods are illustrated by recent examples. The use of additional information from other methods, joint applications of X-ray and neutron scattering, and the possibilities for assessing and validating the models constructed based on small-angle scattering data will be discussed.

1. Introduction

Small-angle scattering (SAS) of X-rays (SAXS) and neutrons (SANS) is applicable to various fields, from metal alloys and polymers to porous materials, nanoparticles, emulsions and biological macromolecules in solution. The main principles of SAXS were developed in the seminal work of A. Guinier (1939) following his studies of metallic alloys. The scattering of X-rays close to the primary beam was found to give structural information on the electron-density fluctuations with dimensions between one and a few hundred nanometres. Since the 1960s, SAXS has been employed for the study of biological macromolecules in solution, where it has provided low-resolution structural information in the absence of crystals. Major improvements of SAXS and SANS instrumentation experiments came in the 1970s with bright synchrotron radiation and steady-state neutron sources. The latter showed the power of contrast variation by H/D exchange (Engelman & Moore, 1972; Ibel & Stuhrmann, 1975). Nowadays, yet brighter possibilities are opened by the third- and forthcoming fourth-generation synchrotrons and advanced spallation neutron sources. Obtaining precise scattering patterns from solutions of biological macromolecules becomes a matter of seconds or less, and, curiously, the limiting steps in time-resolved experiments are now not the data-collection times but rather fast mixing techniques (Akiyama *et al.*, 2002). Given this enormous progress, the major bottleneck in the study of biomacromolecular solutions remains unambiguous interpretation of the data. Modern biological users of SAS cannot be satisfied with just a few overall parameters like the radius of gyration R_g as was the case in the 1960s; instead, interpretation in terms of three-dimensional (3D) models is expected. As dilute solutions of macromolecules yield radially averaged, one-dimensional (1D) scattering patterns, the main challenge of SAS is to extract information about the 3D structure of the object from these 1D experimental data.

Guinier & Fournet (1955) had already demonstrated in their first textbook that SAS yields not just information on the sizes and shapes

but also on the internal structure of particles in various disperse systems. A number of studies have indeed been performed where SAS was able to reconstruct the 3D structure. A striking example of the capabilitiy of SAS to build detailed models of biological macromolecules is provided by the analysis of bacterial virus T7. This large bacteriophage with a molecular mass of about 55 MDa consists of an icosahedral protein capsid with a radius of about 26 nm containing a double-stranded DNA and a cylindrical tail. A SAXS-based structural model of T7 was published by Svergun *et al.* (1982) in a paper presenting an *ab initio* data-interpretation method based on spherical harmonics. The X-ray scattering pattern was well fitted up to a resolution of about 2.5 nm in an axially symmetric approximation to yield not only the overall shape, but also the details of the internal structure (Fig. 1A). The most intriguing was a core with a radius of about 13 nm, suggested to consist of protein, and a cylindrical protrusion observed in the lower part of the virus presumed to be rearranged DNA. Of course, these results could not be further validated at that time by other methods. Very recently, a cryo-electron microscopy (cryo-EM) study has been published of the two phage T7 assemblies produced during its maturation: the DNA-free prohead and the mature virion (Agirrezabala *et al.*, 2005). The former structure (Fig. 1C) visualizes a complex protein assembly in the interior of the capsid, while the latter (Fig. 1B) reveals important changes in the protein shell and in the core complex, which protrudes from the shell to interact with the phage tail. The latter appears thinner in Fig. 1C than in the SAXS model (Fig. 1A), but otherwise the agreement between the structural features predicted by SAXS and the recent cryo-EM reconstruction is excellent. Clearly, the modern and more detailed cryo-EM models (Agirrezabala *et al.*, 2005) provide important new insights towards understanding the macromolecular interactions which take place during the virus assembly and maturation, but one should not forget that the SAXS model was generated nearly 25 years ago.

The last decade brought a long-awaited breakthrough in SAXS/SANS data-analysis methods allowing reliable *ab initio* shape and
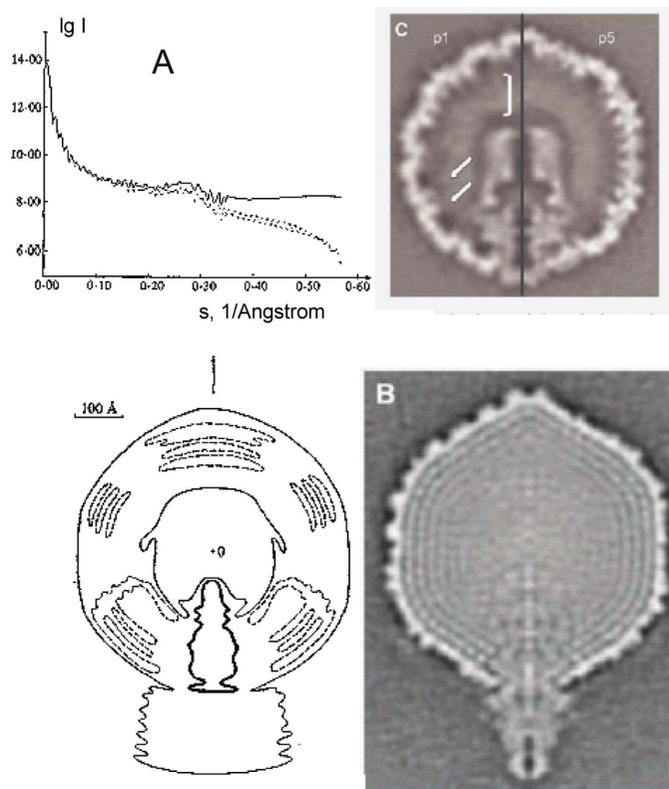
Figure 1
Structure of bacterial virus T7. (A) The model constructed *ab initio* by Svergun *et al.* (1982) from the SAXS data. The processed experimental curve is displayed in the upper panel as a solid line and the scattering from the model is shown as a dashed line. The plot displays the logarithm of the scattering intensity $I(s)$ as a function of the modulus of the scattering vector $s = (4\pi/\lambda)\sin\theta$, where 2θ is the scattering angle and $\lambda = 0.15$ nm is the X-ray wavelength. Bottom panel: a cross section of the electron density map containing the rotation axis. (B) and (C) are cross sections of the cryo-EM 3D structure of the mature virion and of the DNA-free prohead, respectively [adapted with permission from Agirrezabala *et al.* (2005)].

domain structure determination and detailed modelling of macromolecular complexes using rigid-body refinement. The present paper will focus on reviewing these advanced interpretation techniques, which will also be illustrated by some practical applications.

2. *Ab initio* methods

Analysis of monodisperse solutions of biological macromolecules is one of the rare cases when 3D models can be constructed from isotropic SAXS or SANS patterns. The reconstruction is possible thanks to the fact that for dilute monodisperse solutions the solute intensity after solvent subtraction is proportional to the single-particle scattering averaged over all orientations. Of course, *ab initio* analysis is only feasible at a low resolution (1–2 nm) and is thus often limited to homogeneous models (shape determination). In the 1960s, trial-and-error shape modelling was the only available technique, where scattering patterns were computed from different shapes and compared with the experimental data. This situation was changed after Stuhrmann (1970b) introduced the multipole expansion by representing the particle scattering density in spherical coordinates $(r, \omega) = (r, \theta, \varphi)$ as

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

where the spherical harmonics $Y_{lm}(\omega)$ are combinations of trigonometric functions of orders l and m. Here, the truncation value L determines the accuracy of the expansion as the lower-order harmonics define the gross structural features of the particle and the higher harmonics describe finer details. The radial functions are expressed as

$$\rho_{lm}(\mathbf{r}) = \int_{\omega} \rho(\mathbf{r})Y_{lm}^*(\omega) \, d\omega. \qquad (2)$$

The particle scattering amplitude is similarly represented in reciprocal space,

$$A(s) = \sum_{l=0}^{L} \sum_{m=-l}^{l} A_{lm}(s)Y_{lm}(\Omega), \qquad (3)$$

where the partial amplitudes $[A_{lm}(s)]$ are

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

and $j_l(sr)$ are spherical Bessel functions. This leads to a simple expression for the SAS intensity:

$$I(s) = \sum_{l=0}^{L} I_l(s) = 2\pi^2 \sum_{l=0}^{L} \sum_{m=-l}^{l} |A_{lm}(s)|^2, \qquad (5)$$

which is a sum of independent contributions from the substructures corresponding to different spherical harmonics $Y_{lm}(\omega)$. Using the multipole expansion, scattering patterns from known structures can be rapidly computed and the problem of obtaining information about the structure given the SAS intensity can also be meaningfully approached.

The use of spherical harmonics boosted SAS data-analysis methods and this representation is still actively used to construct advanced algorithms. In the first (and very elegant) *ab initio* shape-determination method (Stuhrmann, 1970a) the particle was described by an angular envelope function $F(\omega)$ conveniently expressed *via* $Y_{lm}(\omega)$. This method was further developed (Svergun & Stuhrmann, 1991) and the first publicly available program *SASHA* was written (Svergun *et al.*, 1997). The envelope representation has intrinsic limitations (*e.g.* it is impossible to account for holes inside the particle) but it played a very important role in demonstrating for the first time that under certain circumstances a unique 3D shape can be extracted from the SAS data (Svergun *et al.*, 1996).

Another boost to the development of *ab initio* methods occurred at the end of 1990s, when Monte Carlo type methods became popular (in part, thanks to powerful computers). The idea of the Monte Carlo search in a confined volume was first proposed by Chacon *et al.* (2000, 1998) using a genetic algorithm. A more general approach described below (Svergun, 1999) is also suitable for the analysis of multi-component complexes (*ab initio* shape determination is a particular case of this procedure).

Let us assume that the particle contains K distinct components with different contrasts $\Delta\rho_k$ (*e.g.* a nucleoprotein complex would be a two-component particle with $K = 2$ consisting of the protein and nucleic acid moieties). A volume expected to enclose the particle (*e.g.* a sphere of radius $R = D_{max}/2$, where D_{max} is the maximum size) is filled with densely packed beads of radius $r_0 \ll R$. Each bead is assigned an index X_j indicating the component ('phase') to which it belongs [X_j ranges from 0 (solvent) to K]. The particle structure is described by a configuration vector $\mathbf{X}$ of length $M \simeq (R/r_0)^3$ and the scattering intensity is

$$I(s) = \left\langle \left[\sum_{k=1}^{K} \Delta\rho_k A_k(s) \right]^2 \right\rangle_{\Omega} = \sum_{k=1}^{K} \Delta\rho_k I_k(s) + 2\sum_{n>k} \Delta\rho_k \Delta\rho_n I_{kn}(s), \qquad (6)$$

where $A_k(s)$ and $I_k(s)$ are the scattering amplitude and intensity, respectively, from the volume occupied by the kth phase, and $I_{kn}(s)$ are the cross terms. Spherical harmonics are used to rapidly evaluate the scattering from such a model for a given assignment X_j as described by Svergun (1999).

If a set of $N_C \geq 1$ contrast variation curves $I^j_{\mathrm{exp}}(s)$, $j = 1, \ldots, N_C$ is available, one can search for a configuration X fitting all these curves simultaneously by minimizing the 'energy' function

$$F(X) = \sum_{j=1}^{N_C} \chi_j^2 + \alpha P. \tag{7}$$

Here, the first term contains individual discrepancies between the experimental and calculated curves,

$$\chi_j^2 = \frac{1}{N-1} \sum_{i=1}^{N} \left[\frac{I^j_{\mathrm{exp}}(s_i) - \eta I_{\mathrm{calc}}(s_i)}{\sigma^j(s_i)} \right], \tag{8}$$

where N is the number of experimental points and η is a scaling factor.

As the search models usually contain thousands of beads, the solution is constrained by the second, penalty, term $P(X)$ requiring compactness and connectivity of the individual components in the model (the penalty weight $\alpha > 0$ is selected to ensure a proper balance between the discrepancy and compactness).

The 'energy' minimization is done starting from a random initial approximation by simulated annealing (SA) (Kirkpatrick *et al.*, 1983). At each SA move, the assignment of a single bead is randomly changed and the amplitudes in equations (4)–(6) are updated accounting for this change but not completely recalculated. This accelerates the computations significantly and makes millions of function evaluations required for a typical SA run in acceptable times (depending on the complexity, from a few minutes to a few days on a typical PC).

A multiphase *ab initio* analysis program *MONSA* was used to analyse the neutron contrast variation data from ribosomes (Svergun & Nierhaus, 2000), where simultaneous fitting of 42 scattering patterns from specifically perdeuterated particles allowed one to construct a map of the protein–RNA distribution in the 70S ribosome of *E. coli*. *MONSA* can however also be employed to build *ab initio* models of protein complexes using X-ray scattering. Indeed, in a complex of two proteins A + B one cannot distinguish between protein A and protein B based on a single SAXS curve. However, if the scattering patterns from the individual proteins, A and B, are also available and one can assume that their low-resolution structures do not change in the complex, the latter can be considered as a two-component system with three curves to fit. Such an application of *MONSA* is also possible for multidomain proteins using the scattering patterns recorded from deletion mutants (see the example in §4) (Petoukhov *et al.*, 2006).

For a single curve from a single-component particle (*e.g.* a protein in aqueous solution) the above approach reduces to the *ab initio* shape determination procedure implemented in the program *DAMMIN* (Svergun, 1999). Both *MONSA* and *DAMMIN* permit one to account for *a priori* information about the particle (*e.g.* anisometry, symmetry) (Petoukhov & Svergun, 2003, 2006). *DAMMIN* has been actively used for practical shape determination by different groups (Aparicio *et al.*, 2002; Arndt *et al.*, 2003; Bugs *et al.*, 2004; Dainese *et al.*, 2005; Egea *et al.*, 2001; Fujisawa *et al.*, 2001; Hammel *et al.*, 2004; Scott *et al.*, 2002). Interestingly, the program also provided meaningful models for non-biological systems with moderate polydispersity (Ozerin *et al.*, 2005; Shtykova *et al.*, 2003). The executables of *DAMMIN* and of other programs developed at

the EMBL and described below are freely available for academic users at http://www.embl-hamburg.de/ExternalInfo/Research/Sax/software.html. Other *ab initio* approaches include the original genetic algorithm program *DALAI_GA* (Chacon *et al.*, 2000, 1998; http://sbg.cib.csic.es/Software/Dalai_GA/index.html), a 'give-n-take' program *SAXS3D* (Walther *et al.*, 2000; Bada *et al.*, 2000; http://www.cmpharm.ucsf.edu/~walther/saxs/) and a spheres modelling program *GA_STRUCT* (Heller *et al.*, 2002). Several papers have been published devoted to a comparison of different *ab initio* programs (Takahashi *et al.*, 2003; Zipper & Durchschlag, 2003).

An intrinsic limitation of the *ab initio* shape determination methods is the assumption of uniform scattering length density, which allows one to fit only a restricted portion of the data (usually up to about 2 nm resolution). In an alternative *ab initio* approach (Svergun *et al.*, 2001), the protein is represented by an assembly of dummy residues (DR) instead of beads, whereby each DR has a form factor equal to that of an average residue in water. The method, implemented in the program *GASBOR*, starts from a randomly distributed gas of DRs in a spherical search volume of diameter D_{max}. The number of DRs is usually known *a priori* from the sequence, and the method employs SA to find the coordinates of the DRs fitting the experimental data and building a protein-like structure. For this, the DRs are randomly relocated within the search volume to minimize an energy function of the type in equation (7), where the penalty $P(X)$ requires the model to have a 'chain-compatible' spatial arrangement of DRs. In particular, each DR is required to have two neighbours at a distance of 0.38 nm (the separation between C_α atoms in the polypeptide chain). The use of DRs permits one to fit the experimental data up to a resolution of about 0.5 nm, and *GASBOR* may provide more detailed models than the bead modelling programs, especially for smaller (< 100 kDa) proteins. The DR modelling method is now routinely applied for proteins and protein complexes (Davies *et al.*, 2005; Hough *et al.*, 2004; Shi *et al.*, 2005; Solovyova *et al.*, 2004; Witty *et al.*, 2002).

Given the intrinsic ambiguity of SAS data interpretation, running the Monte Carlo methods from random starts produces somewhat different models yielding nearly identical scattering patterns. These models can be superimposed and averaged to obtain the most probable and an averaged model, which is done automatically in the program package *DAMAVER* (Volkov & Svergun, 2003). Here, a program *SUPCOMB* (Kozin & Svergun, 2001) is repeatedly used to align and compare pairs of models represented by beads or DRs, and this analysis also allows one to assess the reliability of the *ab initio* modelling. Thus, it was reported by Volkov & Svergun (2003) that flat particles starting from anisometry of about 1:5 are difficult to restore *ab initio* but also that for such less reliable reconstructions significant deviations between the individual models are observed. The analysis of multiple solutions is indispensable for practical applications to obtain guidance about the uniqueness of the model. One should stress here that the *ab initio* SAS models are always defined up to an enantiomorphic shape which gives the same scattering curve as the original one. This must also be taken into account *e.g.* when a low-resolution *ab initio* model is compared with the crystal structure of a homologous protein (by default the program *SUPCOMB* makes alignments for enantiomorphs as well). Among other tools to manipulate bead models, the program *SITUS* can be mentioned, which allows one to convert the models into density maps (Wriggers & Chacon, 2001).

An interesting example of an *a posteriori* validation of an *ab initio* model of a protein structure is presented in Fig. 2. The scattering pattern from the C subunit of V-ATPase containing 401 residues (top panel) was fitted using *DAMMIN* at about 2.5 nm resolution and by

GASBOR up to 0.7 nm resolution. The most probable models out of a dozen reconstructions are displayed in the bottom panel (Armbruster *et al.*, 2004). The crystal structure of this protein was independently determined and published later (Drory *et al.*, 2004), and shows a very good agreement with the SAXS-predicted shapes as illustrated by the overlaps in Fig. 2, bottom panel. About 50 residues are still missing in the crystallographic model, which may account for some of the unfilled portions in the SAXS-derived models [note also
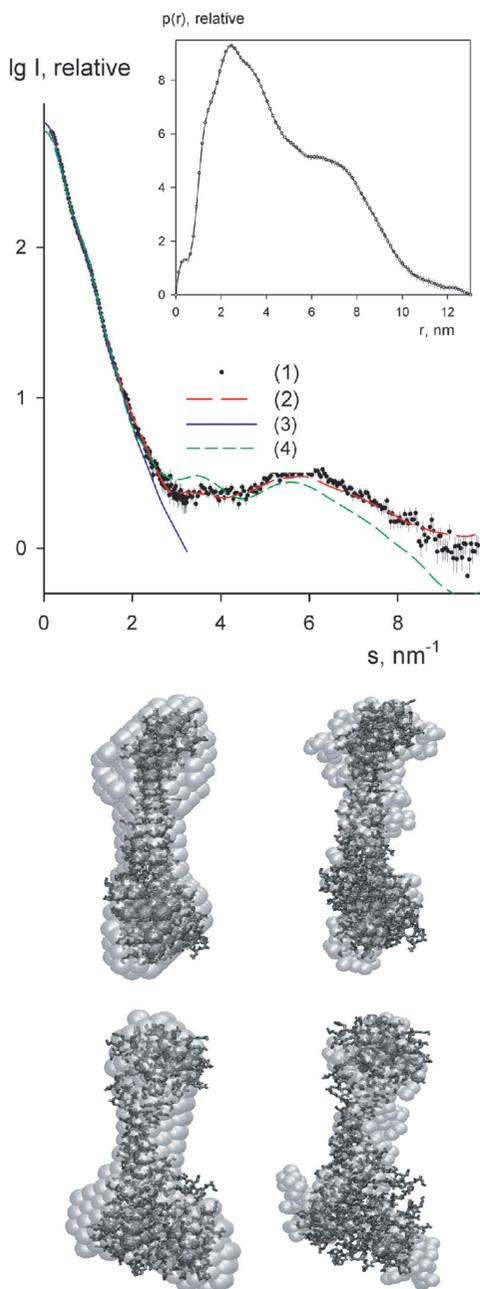
Figure 2
X-ray scattering pattern and models of subunit C from yeast V-ATPase. Top: (1) experimental SAXS curve; (2) and (3) scattering from typical *ab initio GASBOR* and *DAMMIN* models, respectively; (4) scattering from the high-resolution model of subunit C which was determined later [PDB code 1U7L; Drory *et al.* (2004)] calculated by the program *CRYSOL* (Svergun *et al.*, 1995). The computed distance distribution function of subunit C is displayed in the insert. Bottom: *ab initio* low-resolution models reconstructed by *DAMMIN* and *GASBOR* (displayed as beads and dummy residues, respectively), superimposed with the crystallographic model (wire frames). Models in the bottom row are rotated counterclockwise by 90° around the *y* axis.

that the theoretical pattern computed from the crystal structure (curve 4 in the top panel) shows systematic deviations from the experimental data].

3. Rigid-body modelling

Modern 'post-genomic' initiatives aimed at large-scale structure analysis by crystallography and NMR (Gerstein *et al.*, 2003) provided unprecedented amounts of high-resolution structures of individual macromolecules. Obtaining high-resolution models of macro-molecular complexes is often more difficult, but the structures of individual subunits can be successfully employed for rigid-body modelling using lower-resolution methods (cryo-EM or SAXS/SANS). The idea of constructing a model against SAS data by movements and rotations of subunits is not new. Modelling using assemblies of simple geometrical bodies was described *e.g.* by Glatter (1972), and Pavlov (1985) published a method to calculate the scattering from two-domain proteins. Over the years, different approaches have been proposed for rigid-body modelling, where simplified representations of the subunits were often employed. Thus, Wall *et al.* (2000) replaced the subunits by triaxial ellipsoids to find their approximate arrangement followed by docking of the atomic models. In the constrained fit procedure of Boehm *et al.* (1999) and Sun *et al.* (2004), the high-resolution models are reduced to bead assemblies and thousands of possible bead models are screened, also accounting for other results, *e.g.* from ultracentrifugation. In general, the use of information from other methods is extremely valuable for building sound rigid-body models. Useful constraints are obtained about contacting residues from site-directed mutagenesis or from the distances determined by fluorescence studies (Krueger *et al.*, 2000), and from data on surface complementarity and energy minimization (Tung *et al.*, 2002). The use of residual dipolar coupling from NMR helps one to further reduce rotational degrees of freedom during the modelling (Grishaev *et al.*, 2005; Mattinen *et al.*, 2002).

Recent methods using high-resolution structures employ spherical harmonics to accurately compute the scattering from individual subunits and to further rapidly evaluate the scattering from the complex (Petoukhov & Svergun, 2005; Svergun, 1991, 1994). This paves the way for convenient interactive and automated rigid-body modelling. Let us consider a complex consisting of several subunits with known atomic structures. First, the scattering patterns from the subunits must be computed, which can be conveniently done using the programs *CRYSOL* for X-rays (Svergun *et al.*, 1995) and *CRYSON* for neutrons (Svergun *et al.*, 1998). These programs calculate the scattering from the atomic model of the particle in solution as

$$I(s) = \left\langle |A(s)|^2 \right\rangle_{\Omega} = \left\langle |A_{\mathrm{a}}(s) - \rho_{\mathrm{s}}A_{\mathrm{s}}(s) + \delta\rho_{\mathrm{b}}A_{\mathrm{b}}(s)|^2 \right\rangle_{\Omega}, \qquad (9)$$

where $A_{\mathrm{a}}(s)$ is the (X-ray or neutron) scattering amplitude from the particle in vacuum, $A_{\mathrm{s}}(s)$ and $A_{\mathrm{b}}(s)$ are, respectively, the scattering amplitudes from the excluded volume and the hydration shell, both with unit density, and $\langle\ldots\rangle$ denotes the spherical average in reciprocal space. It is taken into account here that the density of the bound solvent ρ_{b} may differ from that of the bulk ρ_{s} such that $\delta\rho_{\mathrm{b}} = \rho_{\mathrm{b}} - \rho_{\mathrm{s}}$ is the contrast of the hydration shell. Given the atomic coordinates, the programs fit the experimental scattering curve by adjusting the excluded volume and the contrast of the hydration layer to minimize the discrepancy in equation (9) (or, in the absence of the experimental data, they predict the theoretical scattering pattern using default or user-defined parameters). The three terms in equation (9) are computed using the multipole expansion as in equations

(3)–(5) to speed up the calculations. *CRYSOL* and *CRYSON* are widely used for validation of crystal structures or theoretical models against SAS data (Hough *et al.*, 2004; King *et al.*, 2005; Vestergaard *et al.*, 2005) but they also provide the partial amplitudes $A_{lm}(s)$, allowing one to rapidly compute the scattering from complexes. Indeed, consider for simplicity a complex of two subunits A and B, where without loss of generality one of the subunits (say, A) can be fixed and the other (B) is moved and rotated during the modelling. The partial amplitudes of the subunits in reference orientations, $A_{lm}(s)$ and $B_{lm}(s)$ are pre-computed using *CRYSOL* or *CRYSON*. If one rotates subunit B by the Euler angles α, β, γ and translates it by a vector $\mathbf{u}$, the scattering intensity of the complex is

$$I(s, \alpha, \beta, \gamma, \mathbf{u}) = I_a(s) + I_b(s) + 4\pi^2 \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \text{Re}\left[A_{lm}(s)C_{lm}^*(s)\right]. \quad (10)$$

Here, $I_a(s)$ and $I_b(s)$ are the scattering intensities from subunits A and B, respectively, which do not change during the modelling, and $C_{lm}(s)$ are the partial amplitudes of the rotated and translated subunit B. The latter can be analytically expressed using the pre-calculated partial amplitudes $B_{lm}(s)$ and the six parameters α, β, γ and $\mathbf{u}$, as described by Svergun (1991, 1994). This approach allows one to rapidly compute the intensity $I(s, \alpha, \beta, \gamma, \mathbf{u})$ for arbitrary rotations
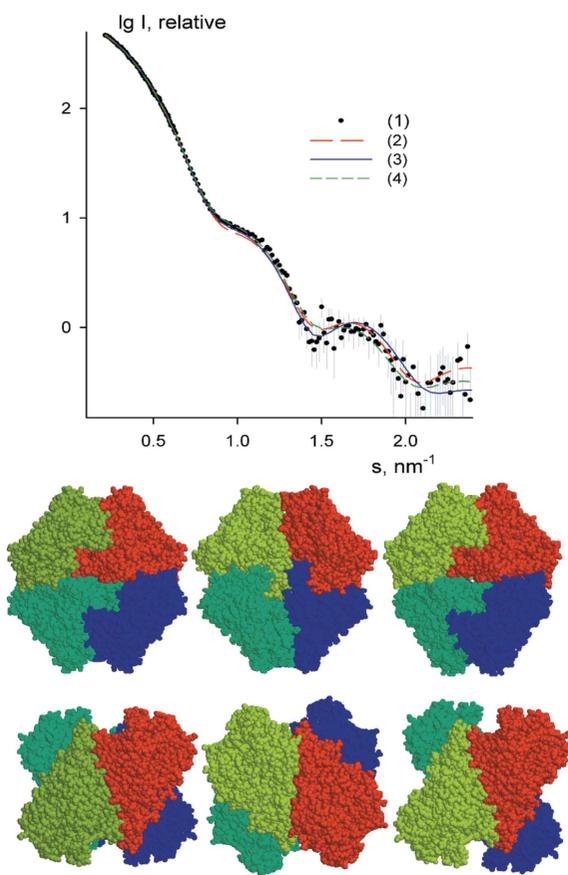
Figure 3
Rigid-body modelling of tetrameric pyruvate oxidase. Top: (1) experimental SAXS curve; (2) computed scattering from the model of POX in the crystal [PDB code 1POW; Muller & Schulz (1993)]; (3) scattering from the model provided by a brute-force method using *GLOBSYMM* with spatial sampling of 0.1 nm and angular sampling of 20°; (4) scattering from the model provided by a global search uisng *SASREF*. Bottom: the models of POX (from left to right: crystallographic structure, *GLOBSYMM* model and *SASREF* model). Models in the bottom row are rotated counterclockwise by 90° around the *x* axis.

and displacements of the second subunit. Equation (10) can be easily generalized for a system of K rigid bodies, which, in the general case, will be described by $6(K-1)$ positional parameters, but this number can be significantly lower *e.g.* for symmetric structures.

There are several approaches utilizing this fast computation of the scattering from macromolecular complexes for rigid-body modelling. Interactive modelling implemented in the programs *ASSA* (Kozin & Svergun, 2000; Kozin *et al.*, 1997) and *MASSHA* (Konarev *et al.*, 2001) permits one to fit the experimental data by manipulating the subunits on the computer display. Such interactive modelling makes it possible to account for additional information (*e.g.* about symmetry, about contacts between subunits *etc.*), and local automated refinement using an exhaustive search in the vicinity of the current configuration is also available. To relieve the users of the burden of interactive analysis, a comprehensive set of tools for automated rigid-body modelling has recently been developed by Petoukhov & Svergun (2005). Depending on the complexity of the system, either exhaustive or heuristic algorithms are employed. Thus, for homo- or hetero-dimeric complexes and for symmetric oligomers with one subunit in the asymmetric part the number of parameters describing the complex is sufficiently low (six at most). The programs *DIMFOM* and *GLOBSYMM* implement different types of 'brute-force' modelling using exhaustive searches for these two cases. The accuracy of the exhaustive search depends on the sampling grid and can be defined by the user. Usually, spatial sampling of about 0.1 nm and angular sampling of about 10° (the default parameters) yield reliable results and can be done within reasonable computational times on a PC (minutes to hours).

For complexes containing several symmetrically unrelated subunits, the conformational space to be explored is too broad for brute-force calculations. A heuristic search is therefore implemented in the program *SASREF*, probably the most versatile program for automated rigid-body modelling (Petoukhov & Svergun, 2005). The idea of the algorithm is rather simple: for a complex consisting of several subunits, one starts from a model with arbitrarily positioned subunits and random rigid-body movements and/or rotations are employed following an SA minimization protocol to find a model fitting the available experimental data. A broad variety of physical and biochemical restrictions can be easily imposed on the search. Appropriate penalty functions are included to ensure that the model is interconnected and displays no steric clashes. Distance restraints between specific residues, loops or entire subunits can be specified to account for information from mutagenesis, footprinting or Fourier transform infrared spectroscopy studies. Orientational restraints from residual dipolar coupling experiments (Grishaev *et al.*, 2005; Mattinen *et al.*, 2002) can also be conveniently incorporated. For symmetric complexes, only the subunits belonging to the asymmetric unit are moved/rotated, and the rest of the complex is generated automatically. If the scattering patterns from subcomplexes are available, these multiple scattering data sets can be simultaneously fitted assuming that the quaternary structure of the subcomplexes remains as in the entire complex. Moreover, valuable information about complexes is provided by contrast variation in SANS, especially when using specific deuteration (Furtado *et al.*, 2004; Heller *et al.*, 2003; King *et al.*, 2005). A possibility for simultaneous fitting of multiple SAXS and SANS contrast variation data sets accounting for specific deuteration of selected subunits has recently been added to *SASREF* (Petoukhov & Svergun, 2006).

The methods mentioned in this section are able to account for the particle symmetry, which provides a very important constraint for the rigid-body modelling. An example is presented in Fig. 3, displaying the analysis of pyruvate oxidase (POX) from *Lactobacillus plan-*

tarum. The experimental SAXS pattern from this tetrameric enzyme is fitted by the scattering curve computed from its crystal structure (PDB code 1POW; Muller & Schulz, 1993) with $\chi = 1.5$, and the fit displays some systematic deviations. In a systematic comparison of the crystal and solution structures of thiamin-dependent enzymes, Svergun *et al.* (2000) analysed the quaternary structure by manual modelling of POX in terms of two dimers. If one uses *GLOBSYMM* for the modelling in terms of the monomer imposing *P*222 symmetry with a rough search grid (angular sampling 20°) the nominally best solution (middle model in the bottom panel in Fig. 3) fits the data better than the crystal structure ($\chi = 1.4$) but breaks the contacts between the neighbouring dimers. In contrast, *SASREF* reproducibly provides the best model (right-hand model in the bottom panel) yielding $\chi = 1.3$ and an r.m.s. of 0.4 nm to the crystal structure. This model is very similar to that proposed by Svergun *et al.* (2000) based on manual modelling. Interestingly, running *GLOBSYMM* using a finer angular sampling (5°) yields nearly the same model as that provided by *SASREF*. However, this example also shows that one must be careful with the interpretation of the rigid-body modelling data as the (incorrect) model obtained with the rough sampling yields a reasonably good fit (better than that from the crystal structure).

In general, accounting for the symmetry is a powerful tool allowing one to construct spectacular models like those of a dimeric complex between hepatocyte growth factor and tyrosine kinase (Gherardi *et al.*, 2006) or of the calcium/calmodulin-dependent protein kinase II holoenzyme with *P*6₂ symmetry (Rosenberg *et al.*, 2005). One should however always keep in mind that SAS alone does not contain explicit information about symmetry and choosing an incorrect symmetry may also yield models compatible with the scattering data. Strong *a priori* evidence about the symmetry and stoichiometry of the complex is therefore required to use the symmetry constraints.

The automated rigid-body refinement methods permit one to construct biochemically sound models of macromolecular complexes. Still, even when accounting for a large body of information, it is possible to obtain multiple solutions providing (nearly) the same fits to the experimental data. Some brute-force analysis programs (*e.g.* *GLOBSYMM*) keep multiple models fitting the data during the search, which are grouped after the minimization is finished, and a list of representative solutions is provided. For Monte Carlo type methods, comparison of the results of several independent reconstructions gives an idea about the stability of the solution, similar to that described above for *ab initio* shape determination. Further assessment and ranking of the results is provided by the analysis of the intersubunit interfaces. A set of simple tools for rapid screening to estimate the quality of the intersubunit contacts using a C_α-only representation and considering shape complementarity and amino-acid composition at the interface is described by Petoukhov & Svergun (2005). More complicated approaches to analysing protein–protein interfaces utilize parameters like solvation potential, residue interface propensity, hydrophobicity, van der Waals interaction, hydrogen bonding and accessible surface area as ranking parameters (*e.g.* Jones & Thornton, 1997; Wang *et al.*, 2002). Methods for 'soft docking' and for combining docking with biochemical and biophysical information have also been developed (Dominguez *et al.*, 2003; Li *et al.*, 2003). A useful tool for the analysis of protein interfaces is the EBI PISA server (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html), which has been developed for the detection of protein assemblies in crystals (Krissinel & Henrick, 2005) but can also be applied to solution models by placing them in sufficiently large unit cells. A careful screening of the assemblies provided by rigid-body modelling methods is often indispensable to resolve or at least reduce the intrinsic ambiguity of SAS-based model building.

4. Combining *ab initio* and rigid-body methods

Multidomain proteins often consist of globular domains connected by linkers. In many practical applications, high-resolution models of the domains are available or can be predicted by homology modelling, whereas the structures of the linkers remain unknown. The linkers may be flexible, which makes the structure analysis of the full-length proteins by crystallography or NMR very difficult. In this case SAS can be used to model the structure by a combined rigid-body and *ab initio* modelling approach. The full-length protein is represented as an assembly of the domains (moved and rotated as rigid bodies) connected by the linkers composed of DRs. In contrast to the freely movable DRs used for *ab initio* methods, the linker is substituted by a flexible chain of interconnected DRs with a spacing of 0.38 nm, and this chain is attached to the appropriate terminal residues of the domains. The X-ray scattering amplitude from such a chain is readily computed (Petoukhov *et al.*, 2002; Svergun *et al.*, 2001) and the scattering from the full-length protein is

$$I(s) = 2\pi^2 \sum_{l=0}^{\infty} \sum_{m=-1}^{l} \left| \sum_k A_{lm}^{(k)}(s) + \sum_j D_{lm}^{(j)}(s) \right|^2, \tag{11}$$

where $A_{lm}^{(k)}(s)$ and $D_{lm}^{(j)}(s)$ are the partial amplitudes of the domains and of the DRs comprising the linkers, respectively.

Like the modelling of multisubunit complexes described above, the search for the optimal model can be done using SA. Starting from an arbitrary configuration of the domains and of the linkers generated as planar zigzag chains, two types of random modifications of the model can be employed. First, a DR belonging to one of the linkers is selected dividing the entire chain into two parts and the smaller part is randomly rotated around this DR. Alternatively, a random rotation is performed of the part of the structure between two randomly selected DRs about the axis connecting these DRs.

The use of equation (11) permits one to easily incorporate multiple data sets from partial constructs into the fitting. In particular, if the experimental data from deletion mutants are available, the scattering from the relevant portions of the model can be computed and all the data can be fitted simultaneously. Like *ab initio* analysis, the 'energy' function has the form

$$E = \sum (\chi^2)_i + \alpha_{\text{cross}} P_{\text{cross}} + \alpha_{\text{ang}} P_{\text{ang}} + \alpha_{\text{dih}} P_{\text{dih}} + \alpha_{\text{ext}} P_{\text{ext}}. \tag{12}$$

Here, the penalty P_{cross} requires the absence of overlaps between the domains and the DR linkers, P_{ang} and P_{dih} are penalties to ensure a proper distribution of bond and dihedral angles, respectively, in the flexible DR chains (Petoukhov *et al.*, 2002), and P_{ext} restricts the radii of gyration of the DR loops (Petoukhov *et al.*, 2002; Petoukhov & Svergun, 2005). The SA protocol searches therefore for a configuration fitting the scattering from the full-length protein (and, optionally, from its shorter constructs) which is free from steric clashes and where the DR loops display native-like conformations.

This algorithm to reconstruct the structure of multidomain proteins against single or multiple scattering data set(s) is implemented in the program *BUNCH* (Petoukhov & Svergun, 2005). Although the method was initially developed for multidomain proteins, it can also be used to generate probable configurations of missing loops in high-resolution structures, as a complementary technique to the methods proposed by Petoukhov *et al.* (2002). One can also employ *BUNCH* for macromolecular complexes consisting of several subunits when the structures of some of the subunits are not known. In this case, not only missing loops within one subunit but also the shape(s) of the missing subunit(s) are restored.

An example illustrating the use of the above method is given by the study of RNA-binding polypyrimidine tract binding protein (PTB). It

contains four RNA recognition motifs (RRMs; each about 10 kDa) connected by flexible linkers, and the high-resolution structures of the four RRMs have previously been determined by NMR (PDB code PTB1; Conte *et al.*, 2000; Simpson *et al.*, 2004). SAXS patterns from full-length PTB and deletion mutants containing all possible sequential combinations of the RRMs were collected (Fig. 4), and all

these constructs were monomeric in solution. The use of *BUNCH* to simultaneously fit all the measured scattering patterns from the PTB constructs (Petoukhov *et al.*, 2006) yielded reproducible results, whereby domains C and D were found to be in close contact, whereas domains B and especially A had loose contacts with the rest of the protein. Interestingly, independent use of the multiphase *ab initio* program *MONSA* (see §2) yielded a consistent low-resolution bead model (see overlap in Fig. 4).

Methods to add missing fragments are now actively used for practical studies (Durand *et al.*, 2006; Garcia *et al.*, 2006; Nardini *et al.*, 2006). It should, however, be kept in mind that the models provided by *BUNCH* reflect average conformations of (often flexible) fragments and can only serve as an indicator of the volume occupied by these fragments, not as a representation of their actual tertiary structure.
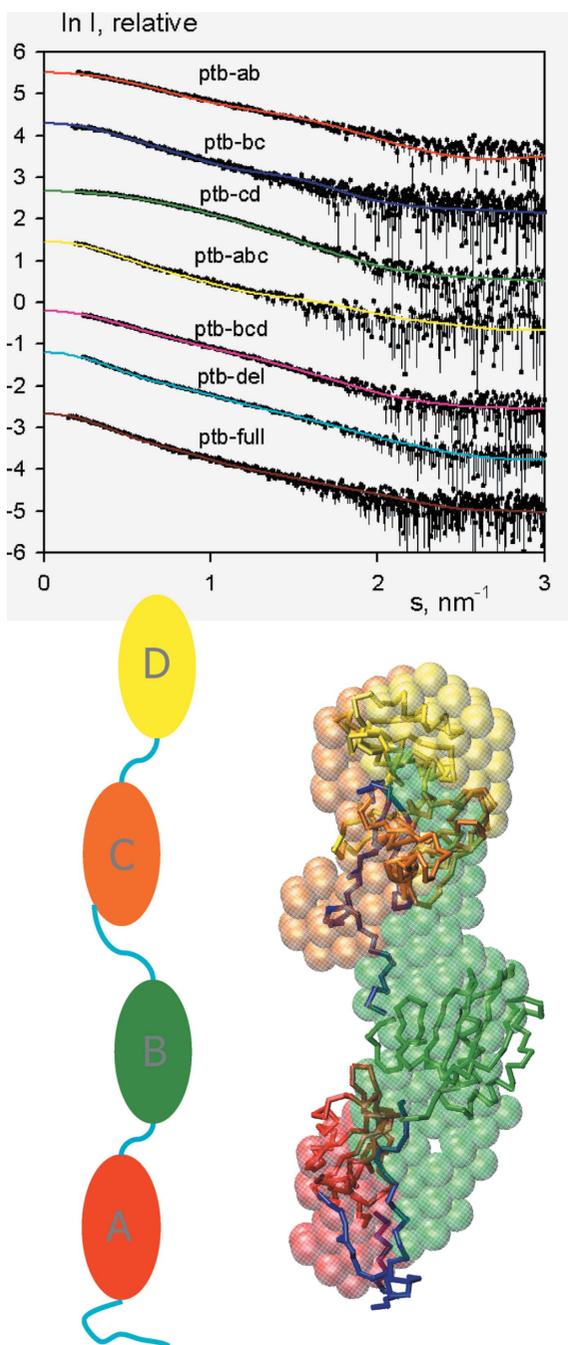
5. Conclusions

Novel approaches to interpret SAS data from solutions of biological macromolecules are especially important given the challenge of the 'post-genomic' era with vast numbers of protein sequences becoming available. Limited by the necessity of growing good-quality crystals for crystallography and by the low-molecular-mass requirement of NMR, most targets for large-scale expression and purification initiatives will probably not be analysed using these two high-resolution methods. SAS should then be the most suitable method to rapidly characterize the targets, at least at low resolution. Even more important is the role of SAS in the analysis of functional macromolecular complexes, as the focus of modern structural genomics is rapidly shifting towards their study. Large assemblies are difficult to analyse by high-resolution methods due to their size, inherent structural flexibility and often transient nature, and rigid-body modelling in SAS, along with cryo-EM, is expected to be at the forefront of this analysis. Large-scale analysis of SAS data will definitely require automation of data acquisition, processing and also interpretation, and steps have already been taken in this direction (Petoukhov *et al.*, 2007). The tremendous progress in SAS instrumentation and novel analysis methods, which has substantially improved the resolution and reliability of the structural models, should make SAS a streamline method for modern structural biology.

Figure 4
Arrangements of domains in polypyrimidine tract binding protein. Top: experimental SAXS curves and the fits computed for the full-length PTB and its deletion mutants. The domain structure of the protein is schematically displayed in the left panel of the bottom figure. The right panel of the bottom figure presents a typical model of PTB generated by *BUNCH* as C$_\alpha$ chains of the NMR structures of individual domains, with colours corresponding to those in the left panel, and linkers shown in blue. Superimposed on the *BUNCH* model is a low-resolution model provided by the multiphase *ab initio* program *MONSA* (semi-transparent beads).

References

Agirrezabala, X., Martin-Benito, J., Caston, J. R., Miranda, R., Valpuesta, J. M. & Carrascosa, J. L. (2005). *EMBO J.* **24**, 3820–3829.

Akiyama, S., Takahashi, S., Kimura, T., Ishimori, K., Morishima, I., Nishikawa, Y. & Fujisawa, T. (2002). *Proc. Natl Acad. Sci. USA*, **99**, 1329–1334.

Aparicio, R., Fischer, H., Scott, D. J., Verschueren, K. H., Kulminskaya, A. A., Eneiskaya, E. V., Neustroev, K. N., Craievich, A. F., Golubev, A. M. & Polikarpov, I. (2002). *Biochemistry*, **41**, 9370–9375.

Armbruster, A., Svergun, D. I., Coskun, U., Juliano, S., Bailer, S. M. & Gruber, G. (2004). *FEBS Lett.* **570**, 119–125.

Arndt, M. H., de Oliveira, C. L., Regis, W. C., Torriani, I. L. & Santoro, M. M. (2003). *Biopolymers*, **69**, 470–479.

Bada, M., Walther, D., Arcangioli, B., Doniach, S. & Delarue, M. (2000). *J. Mol. Biol.* **300**, 563–574.

Boehm, M. K., Woof, J. M., Kerr, M. A. & Perkins, S. J. (1999). *J. Mol. Biol.* **286**, 1421–1447.

Bugs, M. R., Forato, L. A., Bortoleto-Bugs, R. K., Fischer, H., Mascarenhas, Y. P., Ward, R. J. & Colnago, L. A. (2004). *Eur. Biophys. J.* **33**, 335–343.

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.

Conte, M. R., Gröne, T., Ghuman, J., Kelly, G., Ladas, A., Matthews, S. & Curry, S. (2000). *EMBO J.* **15**, 3132–3141.

Dainese, E., Sabatucci, A., van Zadelhoff, G., Angelucci, C. B., Vachette, P., Veldink, G. A., Agro, A. F. & Maccarrone, M. (2005). *J. Mol. Biol.* **349**, 143–152.

Davies, J. M., Tsuruta, H., May, A. P. & Weis, W. I. (2005). *Structure*, **13**, 183–195.

Dominguez, C., Boelens, R. & Bonvin, A. M. (2003). *J. Am. Chem. Soc.* **125**, 1731–1737.

Drory, O., Frolow, F. & Nelson, N. (2004). *EMBO Rep.* **5**, 1148–1152.

Durand, D., Cannella, D., Dubosclard, V., Pebay-Peyroula, E., Vachette, P. & Fieschi, F. (2006). *Biochemistry*, **45**, 7185–9713.

Egea, P. F., Rochel, N., Birck, C., Vachette, P., Timmins, P. A. & Moras, D. (2001). *J. Mol. Biol.* **307**, 557–576.

Engelman, D. M. & Moore, P. B. (1972). *Proc. Natl Acad. Sci. USA*, **69**, 1997–1999.

Fujisawa, T., Kostyukova, A. & Maeda, Y. (2001). *FEBS Lett.* **498**, 67–71.

Furtado, P. B., Whitty, P. W., Robertson, A., Eaton, J. T., Almogren, A., Kerr, M. A., Woof, J. M. & Perkins, S. J. (2004). *J. Mol. Biol.* **338**, 921–941.

Garcia, P., Ucurum, Z., Bucher, R., Svergun, D. I., Huber, T., Lustig, A., Konarev, P. V., Marino, M. & Mayans, O. (2006). *FASEB J.* **20**, 1142–1151.

Gerstein, M., Edwards, A., Arrowsmith, C. H. & Montelione, G. T. (2003). *Science*, **299**, 1663.

Gherardi, E., Sandin, S., Petoukhov, M. V., Finch, J., Youles, M. E., Ofverstedt, L. G., Miguel, R. N., Blundell, T. L., Vande Woude, G. F., Skoglund, U. & Svergun, D. I. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 4046–4051.

Glatter, O. (1972). *Acta Phys. Austriaca*, **36**, 307–315.

Grishaev, A., Wu, J., Trewhella, J. & Bax, A. (2005). *J. Am. Chem. Soc.* **127**, 16621–16628.

Guinier, A. (1939). *Ann. Phys. (Paris)*, **12**, 161–237.

Guinier, A. & Fournet, G. (1955). *Small-angle scattering of X-rays*. New York: Wiley.

Hammel, M., Walther, M., Prassl, R. & Kuhn, H. (2004). *J. Mol. Biol.* **343**, 917–929.

Heller, W. T., Abusamhadneh, E., Finley, N., Rosevear, P. R. & Trewhella, J. (2002). *Biochemistry*, **41**, 15654–15663.

Heller, W. T., Finley, N. L., Dong, W. J., Timmins, P., Cheung, H. C., Rosevear, P. R. & Trewhella, J. (2003). *Biochemistry*, **42**, 7790–7800.

Hough, M. A., Grossmann, J. G., Antonyuk, S. V., Strange, R. W., Doucette, P. A., Rodriguez, J. A., Whitson, L. J., Hart, P. J., Hayward, L. J., Valentine, J. S. & Hasnain, S. S. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 5976–5981.

Ibel, K. & Stuhrmann, H. B. (1975). *J. Mol. Biol.* **93**, 255–265.

Jones, S. & Thornton, J. M. (1997). *J. Mol. Biol.* **272**, 121–132.

King, W. A., Stone, D. B., Timmins, P. A., Narayanan, T., von Brasch, A. A., Mendelson, R. A. & Curmi, P. M. (2005). *J. Mol. Biol.* **345**(4), 797–815.

Kirkpatrick, S., Gelatt, C. D. Jr & Vecci, M. P. (1983). *Science*, **220**, 671–680.

Konarev, P. V., Petoukhov, M. V. & Svergun, D. I. (2001). *J. Appl. Cryst.* **34**, 527–532.

Kozin, M. B. & Svergun, D. I. (2000). *J. Appl. Cryst.* **33**, 775–777.

Kozin, M. B. & Svergun, D. I. (2001). *J. Appl. Cryst.* **34**, 33–41.

Kozin, M. B., Volkov, V. V. & Svergun, D. I. (1997). *J. Appl. Cryst.* **30**, 811–815.

Krissinel, E. & Henrick, K. (2005). *CompLife 2005*, edited by M. R. Berthold, pp. 163–174. Berlin/Heidelberg: Springer-Verlag.

Krueger, J. K., Gallagher, S. C., Wang, C. A. & Trewhella, J. (2000). *Biochemistry*, **39**, 3979–3987.

Li, C. H., Ma, X. H., Chen, W. Z. & Wang, C. X. (2003). *Protein Eng.* **16**, 265–269.

Mattinen, M. L., Paakkonen, K., Ikonen, T., Craven, J., Drakenberg, T., Serimaa, R., Waltho, J. & Annila, A. (2002). *Biophys. J.* **83**, 1177–1183.

Muller, Y. A. & Schulz, G. E. (1993). *Science*, **259**, 965–967.

Nardini, M., Svergun, D., Konarev, P. V., Spano, S., Fasano, M., Bracco, C., Pesce, A., Donadini, A., Cericola, C., Secundo, F., Luini, A., Corda, D. & Bolognesi, M. (2006). *Protein Sci.* **15**, 1042–5100.

Ozerin, A. N., Svergun, D. I., Volkov, V. V., Kuklin, A. I., Gordelyi, V. I., Islamov, A. K., Ozerina, L. A. & Zavorotnyuk, D. S. (2005). *J. Appl. Cryst.* **38**, 996–1003.

Pavlov, M. (1985). *Dokl. Akad. Nauk SSSR*, **281**, 458–462.

Petoukhov, M. V., Eady, N. A., Brown, K. A. & Svergun, D. I. (2002). *Biophys. J.* **83**, 3113–3125.

Petoukhov, M. V., Konarev, P. V., Kikhney, A. G. & Svergun, D. I. (2007). *J. Appl. Cryst.* **40**, s223–s228.

Petoukhov, M. V., Monie, T. P., Allain, F. H., Matthews, S., Curry, S. & Svergun, D. I. (2006). *Structure*, **14**, 1021–1027.

Petoukhov, M. V. & Svergun, D. I. (2003). *J. Appl. Cryst.* **36**, 540–544.

Petoukhov, M. V. & Svergun, D. I. (2005). *Biophys. J.* **89**, 1237–1250.

Petoukhov, M. V. & Svergun, D. I. (2006). *Eur. Biophys. J.* **35**, 567–576.

Rosenberg, O. S., Deindl, S., Sung, R. J., Nairn, A. C. & Kuriyan, J. (2005). *Cell*, **123**, 849–860.

Scott, D. J., Grossmann, J. G., Tame, J. R., Byron, O., Wilson, K. S. & Otto, B. R. (2002). *J. Mol. Biol.* **315**, 1179–1187.

Shi, Y. Y., Hong, X. G. & Wang, C. C. (2005). *J. Biol. Chem.* **280**, 22761–22768.

Shtykova, E. V., Shtykova, E. V Jr, Volkov, V. V., Konarev, P. V., Dembo, A. T., Makhaeva, E. E., Ronova, I. A., Khokhlov, A. R., Reynaers, H. & Svergun, D. I. (2003). *J. Appl. Cryst.* **36**, 669–673.

Simpson, P. J., Monie, T. P., Szendroi, A., Davydova, N., Tyzack, J. K., Conte, M. R., Read, C. M., Cary, P. D., Svergun, D. I., Konarev, P. V., Curry, S. & Matthews, S. (2004). *Structure*, **12**, 1631–1643.

Solovyova, A. S., Nollmann, M., Mitchell, T. J. & Byron, O. (2004). *Biophys. J.*, **87**, 540–52.

Stuhrmann, H. B. (1970a). *Z.. Phys. Chem. Neue Folge*, **72**, 177–198.

Stuhrmann, H. B. (1970b). *Acta Cryst.* A**26**, 297–306.

Sun, Z., Reid, K. B. & Perkins, S. J. (2004). *J. Mol. Biol.* **343**, 1327–1343.

Svergun, D. I. (1991). *J. Appl. Cryst.* **24**, 485–492.

Svergun, D. I. (1994). *Acta Cryst.*, A**50**, 391–402.

Svergun, D. I. (1999). *Biophys. J.* **76**, 2879–2886.

Svergun, D. I., Barberato, C. & Koch, M. H. J. (1995). *J. Appl. Cryst.* **28**, 768–773.

Svergun, D. I., Feigin, L. A. & Schedrin, B. M. (1982). *Acta Cryst.* A**38**, 827–835.

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., Petoukhov, M. V., Koch, M. H. J. & Koenig, S. (2000). *J. Biol. Chem.* **275**, 297–302.

Svergun, D. I., Richard, S., Koch, M. H. J., Sayers, Z., Kuprin, S. & Zaccai, G. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 2267–2272.

Svergun, D. I. & Stuhrmann, H. B. (1991). *Acta Cryst.* A**47**, 736–744.

Svergun, D. I., Volkov, V. V., Kozin, M. B. & Stuhrmann, H. B. (1996). *Acta Cryst.* A**52**, 419–426.

Svergun, D. I., Volkov, V. V., Kozin, M. B., Stuhrmann, H. B., Barberato, C. & Koch, M. H. J. (1997). *J. Appl. Cryst.* **30**, 798–802.

Takahashi, Y., Nishikawa, Y. & Fujisawa, T. (2003). *J. Appl. Cryst.* **36**, 549–552.

Tung, C. S., Walsh, D. A. & Trewhella, J. (2002). *J. Biol. Chem.* **277**, 12423–12431.

Vestergaard, B., Sanyal, S., Roessle, M., Mora, L., Buckingham, R. H., Kastrup, J. S., Gajhede, M., Svergun, D. I. & Ehrenberg, M. (2005). *Mol. Cell*, **20**, 929–938.

Volkov, V. V. & Svergun, D. I. (2003). *J. Appl. Cryst.* **36**, 860–864.

Wall, M. E., Gallagher, S. C. & Trewhella, J. (2000). *Annu. Rev. Phys. Chem.* **51**, 355–380.

Walther, D., Cohen, F. E. & Doniach, S. (2000). *J. Appl. Cryst.* **33**, 350–363.

Wang, R., Lai, L. & Wang, S. (2002). *J. Comput. Aided Mol. Des.* **16**, 11–26.

Witty, M., Sanz, C., Shah, A., Grossmann, J. G., Mizuguchi, K., Perham, R. N. & Luisi, B. (2002). *EMBO J.* **21**, 4207–4218.

Wriggers, W. & Chacon, P. (2001). *J. Appl. Cryst.* **34**, 773–776.

Zipper, P. & Durchschlag, H. (2003). *J. Appl. Cryst.* **36**, 509–514.