ATSAS 2.1, a program package for small-angle scattering data analysis

Petr V. Konarev, Maxim V. Petoukhov, Vladimir V. Volkov and Dmitri I. Svergun

ATSAS 2.1, a program package for small-angle scattering data analysis

Petr V. Konarev,[a,b] Maxim V. Petoukhov,[a,b] Vladimir V. Volkov[b] and Dmitri I. Svergun[a,b]*

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

The program package ATSAS 2.1 for small-angle X-ray and neutron scattering data analysis is presented. The programs included in the package cover the major processing and interpretation steps from primary data reduction to three-dimensional modelling. This system is primarily oriented towards the analysis of biological macromolecules, but could also be used for non-biological isotropic and partially ordered objects (nanoparticle systems, colloidal solutions, polymers in solution and bulk). Recent developments in the programs included in ATSAS 2.1 are highlighted. The main programs run on multiple hardware platforms, including Windows PC, Linux RedHat and Suse, DEC Alpha, SGI IRIX and Mac OSX.

1. Introduction

Small-angle scattering (SAS) of X-rays and neutrons (SAXS and SANS) is widely used to study the low-resolution structure of non-crystalline systems of various nature, including inorganic systems, composite nanomaterials, polymers and biological macromolecules (Feigin & Svergun, 1987). Recent remarkable progress in instrumentation, in particular thanks to high-flux dedicated X-ray synchrotron radiation and neutron beamlines (reviewed *e.g.* by Heenan *et al.*, 1997; Pedersen, 2002) has significantly improved the quality of the experimental SAXS and SANS data. Modern instruments provide enormous amounts of data in high-throughput mode, *e.g.* in time- or space-resolved experiments or measurements under various physical and chemical conditions, such as temperature, pressure, ionic strength, *etc.* Theoretical and methodical developments over the past decade have made it possible to retrieve significantly more structural information from the high-quality SAXS/SANS scattering patterns than previously expected (see *e.g.* Svergun & Koch, 2003, for a review), and the technique is now being employed to address increasingly complicated questions. However, the full value of the structural information in the scattering data is often difficult to realise, in part because of the considerable diversity of often incompatible software tools. Comprehensive analysis of the experimental patterns remains largely a subjective procedure, requiring extensive user intervention. Different large-scale facilities have developed their own procedures and software packages for primary data processing and data analysis to solve particular problems of their user communities. The situation contrasts strongly with *e.g.* protein crystallography, for which very large specialized software packages, like CCP4 (Winn *et al.*, 2002), were developed in a considerable effort covering several decades, and data processing and analysis tools are now largely standardized.

Several useful data processing and manipulation packages created in different laboratories are freely available to the scientific community. Among the processing programs, one may quote for example the packages SANS (Keiderling, 1997) and GRASP (Dewhurst, 2002), the acquisition/reduction system at the DUBBLE ESRF beamline (Homan *et al.*, 2001), an interactive processing program SAXSANA (Hiragi *et al.*, 2003), or the program PRINSAS (Hinde, 2004) for the analysis of porous materials. The program FIT2D (Hammersley, 1995) is a convenient tool for two-dimensional image data reduction/manipulation and peak fitting. Modelling and fitting of one-dimensional curves can be done using the program FISH (Heenan, 1999) for peak analysis and parametric fitting using various form and structure factors, or using the *ab initio* shape determination programs DALAI_GA (Chacon *et al.*, 1998) or SAX3D (Walther *et al.*, 2000).

In the present paper, the program package ATSAS 2.1 is described, which allows the user to perform the major analysis and modelling steps, from basic data processing to enhanced modelling, for one-dimensional scattering data from isotropic systems. The package includes a processing program PRIMUS, which reads different input file formats, and enables data reduction, manipulation and simple fitting. PRIMUS includes interfaces to several programs for the data analysis for monodisperse, polydisperse and interacting systems, mixtures and partially ordered systems. For low-resolution three-dimensional modelling of monodisperse (or moderately poly-disperse) systems, several *ab initio* shape analysis methods are implemented. A rigid-body modelling suite to construct structures of complexes from known high-resolution models of subunits contains both interactive and automated global search programs. Further, ATSAS 2.1 provides utilities for two-dimensional and three-dimensional display, for computation of SAXS/SANS patterns from atomic models, for matching and averaging of three-dimensional models, *etc.* Most of the programs included in the program package are run on multiple hardware platforms, with the exception of PRIMUS and the three-dimensional display program MASSHA, which utilize the graphical user interface (GUI) of Windows-compatible PCs. ATSAS, which is freely available for academic users, allows one to run major data processing and analysis tasks conveniently, and in many cases perform full analysis and modelling of the experimental data from within the same GUI.
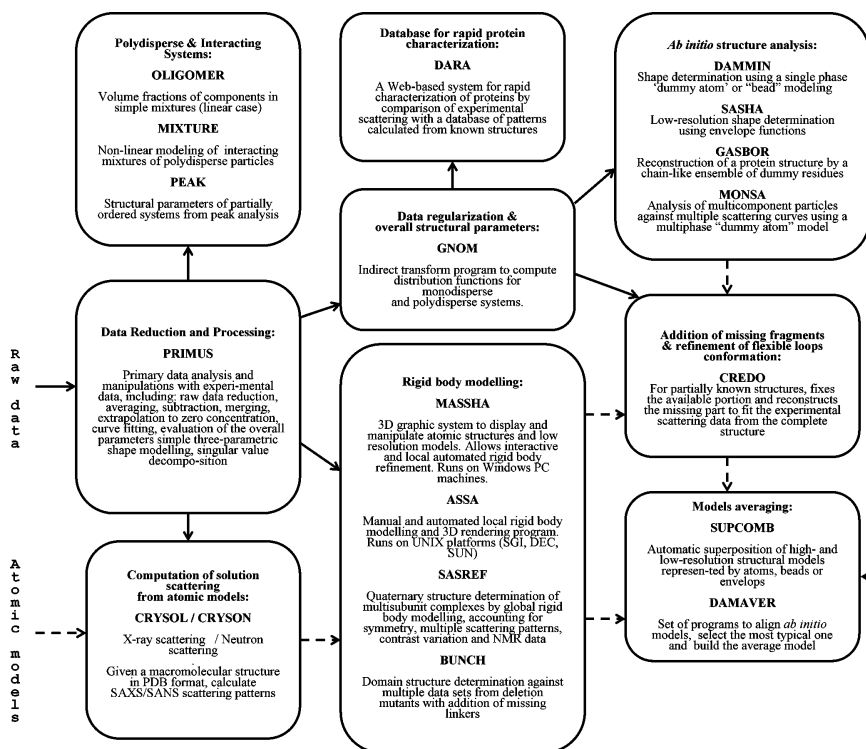
computer programs

Figure 1
A schematic flow chart of *ATSAS* 2.1 programs and their functionality.

The overall structure of *ATSAS* 2.1, with short descriptions of the purpose of individual programs, is presented in Fig. 1. Most of the programs included in *ATSAS* have already been published; in this paper we shall give only a brief account of their functionality (the reader is referred to the original papers for more information). Several programs were however significantly enhanced for release 2.1 of *ATSAS* and these enhancements will be described in more detail.

2. Raw data reduction and processing

The Windows-PC-based program *PRIMUS* (Konarev *et al.*, 2003) is a data processing package encompassing major SAS data analysis steps for isotropic systems, from raw data reduction to simple modelling. This program performs manipulations with one-dimensional data sets and calls other analysis and modelling programs *via* convenient user interfaces. *PRIMUS* contains modules for raw data conversion and processing using the data formats employed at the EMBL SAXS beamline X33 at the storage ring DORIS-III (HASYLAB, DESY, Hamburg), but can also handle reduced data from other facilities, converted by local reduction routines into a generic columnar ASCII format (see §3). Currently, *PRIMUS* allows primary processing of the raw data collected by a linear detector in the OTOKO format (Boulin *et al.*, 1986), and of the two-dimensional data recorded by a MAR345 image-plate detector (http://www.marresearch.com/ip.htm).

The net scattering intensity $I(n)$ as a function of one-dimensional detector channel or averaging bin n (for a two-dimensional detector) is obtained using the standard equation (Strunz *et al.*, 2000)

$$I(n) = \frac{1}{c\,\text{Det}(n)}\left[\frac{I_s(n)}{I_{0s}T_s} - \frac{I_m(n)}{I_{0m}T_m} - \frac{I_e(n)}{I_{0e}}\left(\frac{1}{T_s} - \frac{1}{T_m}\right)\right], \quad (1)$$

where subscripts s, m and e denote the scattering from sample, matrix and empty cell, respectively, subscript 0 refers to the intensity of the incident beam, and T and c stand for sample transmission and the

specimen concentration, respectively. Depending on the type of system, some terms may be missing in equation (1); thus, for $T_s = T_m$ the third term vanishes; if the matrix scattering is not relevant, the second term is absent. The correction for the detector response function $\text{Det}(n)$ may be applied either to one-dimensional data (linear detector) or to the raw data before averaging (area detector).

The reduction and processing of OTOKO files is done by the program *SAPOKO*, invoked from the *PRIMUS* menu. The data are optionally masked, normalized against calibration channels and corrected with respect to the detector response. For the data sets containing multiple frames, these can be averaged or stored individually. A reference frame can be selected, and only the frames that are statistically compatible with the reference frame will be included in the average (usually, for synchrotron studies the first frame is chosen as the reference to monitor radiation damage). The associated errors are computed using Poisson statistics from the numbers of counts in the raw data (Bevington, 1969). Patterns from standard samples with appropriate periodicities (*e.g.* dry collagen or silver behenate) are employed to generate the angular axis. Using the axis information, the processed binary data are converted from OTOKO to a columnar ASCII format. The ASCII file contains, after the header, the momentum transfer axis 's' in the first column, the intensity '$I(s)$' in the second, and optionally the standard error '$\sigma(s)$' in the third column. The momentum transfer is $s = 4\pi\sin\theta/\lambda$, where 2θ is the scattering angle and λ is the radiation wavelength.

A similar user interface is available for the reduction and radial averaging of the raw data recorded by a MAR image-plate scanner. The program *FIT2D* (Hammersley, 1995) is employed to evaluate the
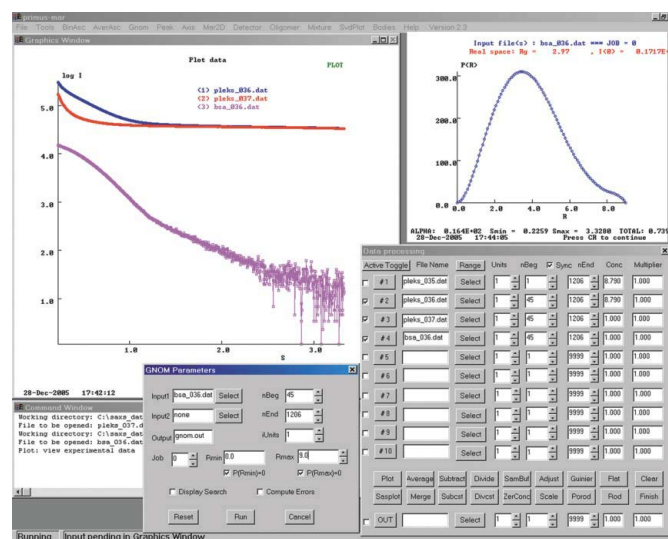

Figure 2
User interface of *PRIMUS* displaying the data processing toolbox, graphics and text window. Blue, red and magenta curves are the scattering profile from a solution of bovine serum albumin, from the solvent and the net scattering after background subtraction, respectively. A 'pop-up' dialog box calling the indirect transformation program *GNOM* and the computed distance distribution function are also displayed.

beam centre position and to generate a mask file. The binary MAR data are normalized by the detector response and the pixels are radially averaged into the appropriate angular bins. The angular axis file for the binning is generated form the scattering patterns of standard samples (typically, silver behenate). The readings of the calibration channels are extracted from the MAR file header and the experimental data are normalized by the transmitted beam intensity and collection time. The output ASCII file has a format similar to that of processed OTOKO files, but with an additional column containing the standard deviation due to the radial average. The averaging algorithm can be easily extended for different formats of the experimental data, and it is planned to add options for raw data processing of other detector types employed in SAXS/SANS studies.

3. Data manipulations

Further data manipulation operations in *PRIMUS*, made with the columnar ASCII files, are independent of the initial raw data format. All the manipulation modules have a smart reader to recognize the beginning of the data stream in the input ACSII file, so that *e.g.* ILL- (Ghosh, 1989), LOQ- (King & Heenan, 1996) or sasCIF- (Malfois & Svergun, 2000) formatted files are read without modification. The user interface of *PRIMUS*, displaying the data manipulation toolbox, is presented in Fig. 2. The angular units in the input files can be converted between nm^{-1} and Å^{-1} and between $s = 4\pi \sin\theta/\lambda$ and $S = s/2\pi$ by a selection bar, and each file can be made active for plotting and other operations by ticking the appropriate box. The range(s) of data points can be adjusted for each data set individually or for selected groups of files simultaneously. A standard plotting operation displays currently active files in semi-logarithmic scale (Fig. 2, left-hand panel). An advanced plotting option invokes the program *SASPLOT*, which displays the files in a separate resizable window using different scales [*e.g.* a fractal (log–log) plot, Porod plot, *etc.*] and permits one to select the plotting range and zoom factor with the mouse. The specimen concentration and the scale multiplier are equal to unity by default, but can be specified by the user (the concentration is automatically read from the header string provided that the latter contains a token 'c = ' followed by a numeric value).

The data manipulation toolbox includes basic arithmetic operations with numbers and two or several data sets (subtraction, division, averaging, *etc.*). Operations with two files also include an adjustment function, which tests the hypothesis that the two data sets are statistically different up to a scale factor and (optionally) a constant term. Several data files can be scaled and optionally spliced and merged using a least-squares fitting in the overlapping range. For this, the experimental data are rebinned onto a new angular master grid and, if requested, the merged curve is recomputed on this grid. This option is useful *e.g.* for making composite curves from the data recorded in different angular ranges. For all operations, the error propagation is performed using standard equations (Bevington, 1969).

The data manipulation dialogue contains ten input file boxes and an output box, where the result of each operation is placed and can be further used in subsequent data manipulations. Information about the operation performed (type of operation, file names, weights, ranges of points used, *etc.*) is written in the header and footer of the output file, and the headers of the operands are appended to the output data stream. The headers and footers of the data files can be visualized in the *PRIMUS* text window to monitor the flow of operations.

Two options are available for data extrapolation to zero specimen concentration in the sample. One option accounts for the finite volume fraction of the specimen, v_k, by solving the system of linear equations

$$I_{\text{exp}}^{(k)}(s) = v_k I_{\text{sample}}(s) + (1 - v_k) I_{\text{matrix}}(s), \qquad (2)$$

where the index k runs over the available measurements in the concentration series to separate the two functions, $I_{\text{sample}}(s)$ and $I_{\text{matrix}}(s)$ (this is useful for subtracting background for protein solution measurements at high angles for high solute concentrations). Another option accounts for attractive or repulsive interactions, which mostly influence the data at very small angles. For this, several scattering patterns at different concentrations are appropriately scaled and the initial portion of the data is extrapolated to zero concentration, assuming linear concentration dependence.

4. Data analysis with *PRIMUS*

4.1. Computation of overall parameters

PRIMUS provides a convenient interface for the calculation of the overall parameters of monodisperse and polydisperse systems from the scattering data. The radius of gyration R_g and forward scattering $I(0)$ are computed using the Guinier approximation:

$$I(s) = I(0) \exp(-s^2 R_g^2/3), \qquad (3)$$

which is valid for $(sR_g) < 1.3$ (Guinier, 1939). The first and last points in the fit can be changed interactively and the Guinier plot with the residuals is automatically redrawn. Similar tools are available to compute the radii of gyration of the thickness and of the cross section for flat and rod-like particles, respectively.

The Porod invariant Q and excluded volume V of the particle (Porod, 1982) are evaluated as

$$V = 2\pi^2 I(0)/Q = 2\pi^2 I(0) \Big/ \int_0^\infty s^2 [I(s) - K] \, ds, \qquad (4)$$

with empirical corrections to compensate for the termination effects (Feigin & Svergun, 1987). The constant K (required to enforce the s^{-4} decay of the intensity) is automatically calculated from the Porod asymptote [$I(s)s^4$ *versus* s^4] at higher angles.

4.2. Calling external data analysis modules

More complicated operations to compute structural characteristics are performed by invoking external program modules from the GUI of *PRIMUS*. In all cases when an external program is called, the file from the output box (or, if the latter is empty, the first active file from the input box) of the data manipulation toolbox is transferred to this program for processing. To assess the overall particle shape rapidly, the program *BODIES* fits the initial portion of the scattering data by the scattering from three-parametric geometrical bodies (triaxial ellipsoid, ellipsoid of revolution, circular, elliptical or hollow cylinder, prism). The scattering intensities from the bodies, $I(s)$, are evaluated using standard formulae (Feigin & Svergun, 1987) and the program minimizes the discrepancy

$$\chi^2 = \frac{1}{N-1} \sum_{j=1}^{N} \left[\frac{\mu I(s_j) - I_{\text{exp}}(s_j)}{\sigma(s_j)} \right]^2, \qquad (5)$$

where N is the number of experimental points and μ is the scaling factor. The best approximation in each class of bodies is found and the appropriate fit(s), along with the parameters, are given.

computer programs

To compute the characteristic functions of dilute monodisperse or polydisperse systems, *PRIMUS* calls the indirect transformation program *GNOM* (Svergun *et al.*, 1988; Semenyuk & Svergun, 1991; Svergun, 1992), which employs the indirect transformation method (Glatter, 1977) to solve the integral equation

$$I_{\exp}(s) = \int_{r_{\min}}^{r_{\max}} p(r) K(s, r) \, \mathrm{d}r, \qquad (6)$$

where $p(r)$ is a distribution function (*e.g.* distance distribution or size distribution) defined in the range $[r_{\min}, r_{\max}]$, and $K(s, r)$ is the integral kernel of the corresponding Fourier transformation. The user-defined parameters, including the type of system (monodisperse or polydisperse), ranges in real and reciprocal space, *etc.*, are specified in the dialog box (Fig. 2). The perception criteria for selecting the regularization multiplier incorporated in *GNOM* (Svergun, 1992) usually provide the distribution function without further user intervention.

Structural characteristics of partially ordered systems are computed from the maxima in the scattering profiles using the program *PEAK*. A single or multiple peak(s) is selected by mouse and fitted by Gaussian functions, allowing also for two-parametric background subtraction. The structural parameters include the periodicity, the long-range order dimension and the degree of disorder, calculated from the peak position and width using standard equations (Vainshtein, 1966).

4.3. Analysis of mixtures: linear and non-linear cases

Several programs are provided for quantitative analysis of the scattering from systems containing distinct components such that the scattering pattern is a linear combination

$$I(s) = \sum_{k=1}^{K} v_k I_k(s), \qquad (7)$$

where n_k and $I_k(s)$ are the volume fraction and the scattering intensity from the *j*th component, respectively. When neither the number nor the intensities of the components are known, model-independent

Figure 3
Experimental scattering pattern from tricorn protein (blue circles) and best fit (red curve) obtained for a mixture of tricorn monomers, dimers and hexamers computed by *OLIGOMER*. On the right-hand side, the intensities from these components are shown. The 'pop-up' menu for calling *OLIGOMER* from *PRIMUS* is displayed at the bottom.

analysis of multiple scattering data sets recorded from the system with varying volume fractions is performed using singular value decomposition (SVD; Golub & Reinsch, 1970). The program *SVDPLOT* computes the SVD of the active data sets in the toolbox to yield the singular vectors (fictitious scattering curves providing an orthogonal basis for linear representation of the entire set of data) and their associated singular values (weights in this basis). The number of non-random singular vectors with significant singular values [evaluated using a non-parametric randomness test (Larson, 1975)] yields the minimum number of independent curves required to represent the entire data set, *i.e.* the number of significant components in the mixture [*e.g.* number of intermediates in a (dis)assembly process].

If the number of components and their scattering intensities are known, the program *OLIGOMER* implements a non-negative linear least-squares algorithm (Lawson & Hanson, 1974) to find the volume fractions n_k, minimizing discrepancy in equation (7). *OLIGOMER* can be launched to process multiple experimental data sets against the same set of functions $I_k(s)$, which is useful *e.g.* in rapid characterization of titration series of oligomeric equilibrium mixtures of proteins with known high-resolution structure. As an example of invoking *OLIGOMER* from *PRIMUS*, Fig. 3 displays SAXS data modelling of a solution of *Thermoplasma acidophilum* tricorn protease. This protein is hexameric in the crystal [PDB entry 1K32 (Brandstetter *et al.*, 2001)] but appears to dissociate partially in solution. The scattering patterns from monomer, dimer and hexamer constructs of this protein were computed by the program *CRYSOL* (described in §6) and the experimental data were fitted by a linear combination of the three components yielding the volume fractions of the three components that provide the best fit to the data (0.55, 0.10 and 0.35 for monomers, dimers and hexamers, respectively).

For more complicated mixtures of different types of particles with possible polydispersity and interparticle interactions, the scattering intensity from each component can be represented as

$$I_k(s) = S_k(s) \int_0^{\infty} D_k(R) V_k(R) [\Delta \rho_k(R)]^2 i_{0k}(s, R) \, \mathrm{d}R, \qquad (8)$$

where $\Delta \rho_k(R)$, $V_k(R)$ and $i_{0k}(s, R)$ denote the contrast, volume and normalized scattering intensity (form factor) respectively of the particle with size R [these functions are defined by the shape and internal structure of the particles, and $i_{0k}(0, R) = 1$], whereas $S_k(s)$ is the structure factor describing the interference effects for the *k*th component.

The program *MIXTURE* employs equations (7) and (8) to characterize quantitatively mixtures of particles with simple geometrical shapes containing up to ten different components (types of particles). Each component is described by its volume fraction, form factor, contrast, polydispersity and, for spherical particles, potential of interparticle interactions (Svergun *et al.*, 2000). Currently, solid or hollow spherical shells or circular cylinders, ellipsoids and dumbbells are supported, for which the form factor is represented by a few parameters (*e.g.* spherical shell radii R_k). The size polydispersity $D_k(R)$ is described by a monomodal distribution characterized by the average dimension R_{0k} and dispersion ΔR_k. Interparticle interactions for spherical particles are accounted by a structure factor $S_k(s)$ in the Perkus–Yevick approximation using the sticky hard-sphere potential (Baxter, 1968).

MIXTURE is run from *PRIMUS* by a 'pop-up' menu prompting for the experimental data file names and a command file name containing specifications of the model and initial values of parameters, *i.e.* number of components, type of each component, relative volume fraction, dimension parameters, average sphere radius, its
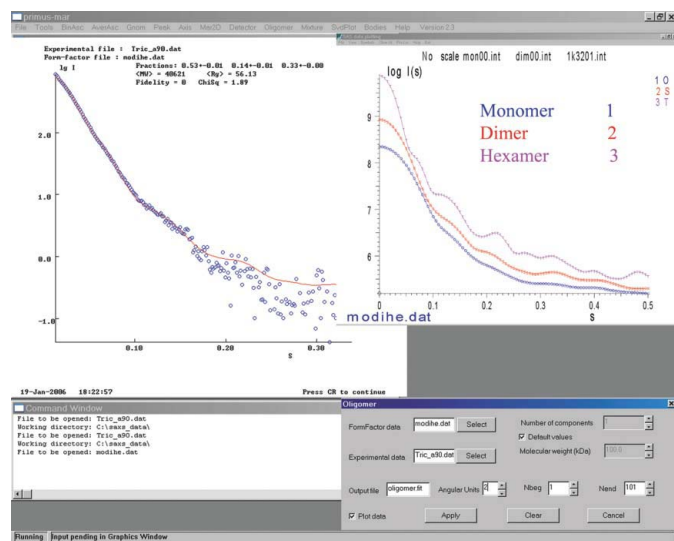
polydispersity, *etc*. If known, upper and lower limits can be specified for all fitting parameters. The final fit to the experimental data is displayed and the partial scattering intensities from the components and the relevant structure factors are stored as separate files. The initial approximation and the best fit parameters for each component are written to a log file after each run to build the history of running *MIXTURE* under different conditions.

5. *Ab initio* shape analysis

For systems of randomly oriented identical particles without interactions (monodisperse systems, *e.g.* dilute solutions of biological macromolecules), SAS provides the possibility to restore the low-resolution shape of particles *ab initio*. A number of methods have been developed employing different shape representations: angular envelope function (Stuhrmann, 1970; Svergun *et al.*, 1996), collections of beads (Chacon *et al.*, 1998, 2000; Svergun, 1999; Walther *et al.*, 2000), interconnected ellipsoids (Vigil *et al.*, 2001; Heller *et al.*, 2002), or, for proteins, dummy residues (Svergun *et al.*, 2001). Below we shall present the new features and improvements in the *ab initio* programs which belong to the *ATSAS* 2.1 package.

5.1. Bead modelling

The *ab initio* shape determination program *DAMMIN* (Svergun, 1999) represents the particle as a collection of a large number of densely packed beads inside a search volume (*e.g.* a sphere with the diameter D_{max} equal to the maximum size of the particle). Each bead belongs either to the particle or to the solvent, and the shape is described by a binary string of length M. The scattering intensity from such a model is rapidly calculated using spherical harmonics as

$$I(s) = 2\pi^2 \sum_{l=0}^{\infty} \sum_{m=-l}^{l} |A_{lm}(s)|^2. \tag{9}$$

Here, the partial amplitudes $A_{lm}(s)$ are

$$A_{lm}(s) = i^l (2/\pi)^{1/2} g(s) \sum_j j_l(sr_j) Y_{lm}^*(\omega_j), \tag{10}$$

where $g(s)$ is the form factor of a single bead, the sum runs over the beads with polar coordinates $(r_j w_j)$ belonging to the particle, $j_l(sr)$ are spherical Bessel functions and $Y_{lm}(w)$ are spherical harmonics.

Starting from an arbitrary string, a Monte-Carlo-type search can be applied by randomly changing the bead assignments to find a configuration that fits the experimental data. Such a search in a confined volume was first proposed by Chacon *et al.* (1998) who used a genetic algorithm (program *DALAI_GA*). In *DAMMIN*, a simulated annealing (SA) protocol is employed to search for a compact model that fits the 'shape scattering' obtained after subtraction of an appropriate constant from the experimental data to force the s^{-4} decay of the intensity at higher angles following Porod's law (Porod, 1982) for homogeneous particles [see equation (4)].

DAMMIN has been extensively developed during recent years following feedback and requests from the user community. In particular, *DAMMIN* allows one to account for various types of particle symmetry. The symmetry is introduced as a rigid constraint, *i.e.* the initial generation, randomization and each modification of the model during the SA procedure are performed for groups of symmetry-related beads simultaneously. The presence of symmetry yields selection rules to spherical harmonics, which are employed to compute the scattering of the bead model in *DAMMIN*. These restrictions not only enable one to obtain more adequate models of symmetric particles, but also significantly speed up the calculations.

The number of symmetry groups in *DAMMIN* is constantly being widened; the program currently supports *Pn* point groups up to *n* = 19-fold rotation symmetry axis, *Pn*2 groups up to *n* = 12 and cubic groups *P*23 and *P*432. All these groups are either compatible with the dense hexagonal packing of beads or they can be implemented by generating beads on appropriate layer lines in cylindrical shells. Most recently, icosahedral symmetry was implemented, which requires a more complicated initial generation. A grid of beads is generated on a set of concentric icosahedral shells, each shell consisting of 20 regular triangles filled by beads on a plane hexagonal lattice, where the number of beads belonging to each edge of the triangle equals the shell's ordinal number. The entire shell is generated as follows: the first triangle is positioned so that one of its vertices belongs to the *Z* axis, the line connecting the other two is parallel to the *XY* plane, and the centre of the triangle has the polar angles $\theta_3 = \arccos\{[5 + 2(5^{1/2})]/15\}^{1/2} \simeq 37.38°$, $\varphi = 36°$. The second triangle is obtained from the first one by a twofold symmetry rotation about the axis connecting the origin with the centre of the bottom edge of the first triangle; corresponding polar angles are $\theta_2 = \arccos\{[5 - (5^{1/2})]/10\}^{1/2} \simeq 58.28°$, $\varphi = 36°$. All other coordinates are generated from these two triangles by a fivefold rotation about *Z* followed by a twofold rotation about *Y*. The separation between the shells corresponds to the distances between the neighbouring beads along the *Z* axis, equal to a double bead radius. Each bead that is not located at a special position has 59 symmetry mates; those belonging to one of the six fivefold axes, or ten threefold axes, or fifteen twofold axes have 11, 19 or 29 symmetry mates, respectively; the central bead has no mates. The icosahedral symmetry implemented in *DAMMIN* should be especially useful in the analysis of virus particles.

If *a priori* information about the particle shape is available, it can also be taken into account. In particular, when overall particle anisometry is known, the requirement of prolateness or oblateness can be added as a soft restraint. Further, the program is able to perform a search within non-spherical search volumes (ellipsoid, cylinder, hollow cylinder or parallelepiped volumes can be used). For the non-spherical volumes, the user is prompted to specify up to three parameters describing the selected geometrical body. For significantly anisometric particles, such a search may improve stability and save computational time.

As with other Monte-Carlo-type methods, different models can be obtained when running *DAMMIN* starting from different initial approximations. To validate the results of the modelling, the program is often run several times and the results of separate reconstructions are compared with one another and averaged. The repetitive runs may be time-consuming, especially for anisometric particles when a large number of beads is required for adequate shape representation. A special 'Keep' mode has thus been included in *DAMMIN* to generate several distinct models in one single run. In contrast to normal modes of operation, where only the best solution is kept during the SA procedure, the 'Keep' mode monitors a set of distinctly different shapes that provide best fits to the experimental data. Further, a modified annealing scheme is employed where the temperature is rapidly decreased and then again increased repetitively. After the minimization, up to 15 models differing from each other by more than 50% of bead positions are saved. These models usually still have similar overall appearance and are ready for comparison and averaging using the program *DAMAVER* (described in §5.4). The other *DAMMIN* modes, listed in order of increasing complexity and required CPU time, are 'Fast', 'Slow' and 'Expert' (the latter allowing the advanced user to tune the SA parameters).

The three modes typically require about 1, 10 and 24 h of computing time on a 1.5 GHz Windows PC; the 'Keep' mode is close to the 'Slow' mode in performance.

During the SA procedure, *DAMMIN* does not fit the experimental data directly, but uses a regularized scattering curve $I_{GNOM}(s)$ provided by the program *GNOM* (see §4.2). In many cases, the preprocessing by *GNOM* permits one to reduce unwanted interference effects in the scattering data and to define the D_{max} value better. Further, to speed up the calculations, the number of knots where the scattering from the bead model is computed can be significantly reduced. Indeed, following the Shannon sampling theorem (Shannon & Weaver, 1949), the scattering intensity can be represented by its values in a discrete set of points $s_k = k\pi/D_{max}$ and interpolated between them. Model calculations indicate that fitting of the preprocessed *GNOM* curve at two knots per Shannon channel instead of fitting every point in the experimental curve does not compromise the convergence of the method. The experimental SAS data (especially SAXS) are usually significantly oversampled, and this reduction accelerates the computations by a factor of ten or more. Another advantage of *GNOM* preprocessing is the ability to take into account the instrumental smearing, if the latter is present. This is done by *GNOM* together with the computation of the $p(r)$ function, and the curve $I_{GNOM}(s)$ evaluated by the appropriate Fourier transformation of $p(r)$ (and fitted by *DAMMIN*) is free from instrumental distortions. To compute the final fit to the experimental data, the scattering from the model is interpolated to the experimental grid using cubic splines (Press *et al.*, 1992). If the smearing effects are present, the final fit is additionally transformed using the ratio $I_{smeared}(s) = I(s)I_{Gsmear}(s)/I_{GNOM}(s)$, where $I_{Gsmear}(s)$ is the smeared *GNOM* fit to the experimental data, also read in from the *GNOM* file.

Initially, *DAMMIN* was a dialogue program (although asking only very few questions, except when started in the 'expert' mode). Given its more and more frequent use as a routine tool for data analysis, an option was added to run *DAMMIN* in a batch mode, with the most important parameters listed in the UNIX-like style on the command line. The user must give the name of the *GNOM* file from where to read the data, and may specify the mode of calculation, particle symmetry, *etc.* Any parameters that are not given in the command line are taken at their default values. In the simplest case, only the *GNOM* file name is given and the program determines the shape in 'Fast' mode without symmetry restrictions. A typical command has the following format:

DAMMIN *<name>* /MO *<mode>* /LO *<log file name>*

/SY *<symmetry>* /ID *<project description>*,

where *<name>* is the name of the *GNOM* output file (compulsory; the other parameters are optional); /MO is the mode key, *<mode>* = Fast (default), Slow or Keep; /LO is the log file key (the default value is the prefix of the *GNOM* file name); /SY is the symmetry key (P1 is the default value); /ID allows the user to specify the description of the project; by default it is the content of the command line. Calling *DAMMIN* without parameters starts it in the dialogue mode.

5.2. Dummy-residues modelling

Another approach for *ab initio* domain structure determination of proteins from the SAXS data utilizes the fact that proteins are polypeptide chains composed of amino acid residues separated by approximately 0.38 nm between adjacent C_α atoms in the primary sequence. Up to a resolution of about 0.5 nm, the protein structure can be considered as an assembly of identical dummy residues (DR)

centred at the C_α positions. A three-dimensional model of the protein may be constructed from the SAXS data by finding a chain-like spatial arrangement of the DRs that fits the experimental scattering pattern, including small- and medium-angles parts. The use of the C_α positions permits one to impose restrictions on the spatial arrangement of the DRs. In addition to the 0.38 nm separation along the chain, excluded-volume effects and local interactions lead to a characteristic distribution of nearest neighbours. The program *GASBOR* (Svergun *et al.*, 2001) uses SA to build a locally 'chain-compatible' DR model inside the spherical search volume defined as in *DAMMIN*. The DR modelling is able to account for the internal structure better, to fit higher resolution data and generally provides more detailed models than those given by the shape determination algorithm *DAMMIN*. Besides the original version of *GASBOR*, fitting the data in reciprocal space, there is a real-space version (Petoukhov & Svergun, 2003), which fits the distance distribution function $p(r)$ provided by *GNOM* from the SAXS experimental data.

Similarly to *DAMMIN*, *GASBOR* is able to account for symmetry and anisometry. During minimization in reciprocal space, it computes the fit, like *DAMMIN*, in two points per Shannon channel, and the final output is obtained using spline interpolation followed by smearing as described in the previous section (§5.1.

A command-line mode of *GASBOR* was recently introduced for both reciprocal- and real-space versions. The required parameters are specified along with optional key values and the program runs in the batch mode. The command line in this case has the form:

GASBOR *<name>* N_{DR} /LO *<log file name>*

/SY *<symmetry>* /ID *<project description>*,

where *<name>* is the compulsory name of the *GNOM* output file, N_{DR} is the number of dummy residues in the asymmetric part (compulsory), and the other terms have the same meaning as in *DAMMIN*.

5.3. Addition of missing portions to high-resolution protein models

The 'dummy residues' approach was further extended to reconstruct missing domains in multidomain proteins and to find probable configurations of disordered loops in crystallographic models by Petoukhov *et al.* (2002). The main idea consists of fixing the known part of the structure (either high- or low-resolution model) and modelling the missing portions, such as disordered loops or domains, to fit the experimental scattering data obtained from the entire particle. Where applicable, information about the primary and secondary structure can be used to restrain the model and to provide native-like conformations of the missing structural fragments. Four different computer programs implemented in the *CREDO* package (Petoukhov *et al.*, 2002; Petoukhov & Svergun, 2003) provide tools for various situations in which a structure lacks a loop or a domain. These methods permit the effective use of SAXS for complementing results obtained by high-resolution methods like X-ray diffraction and NMR spectroscopy.

5.4. Automated analysis of independent reconstructions

The program suite *DAMAVER* (Volkov & Svergun, 2003) estimates the stability of the solution provided by *ab initio* modelling programs. It superimposes multiple models generated by *DAMMIN* or *GASBOR*, selects the most probable one and calculates the average model. *DAMAVER* is based on the program *SUPCOMB*, which aligns two arbitrary low- or high-resolution models by minimizing a normalized spatial discrepancy, NSD (Kozin & Svergun, 2001). All pairs of models are compared and the model with the

lowest mean value of NSD is selected as a reference. The models with the mean NSD value exceeding twice the dispersion of NSD are considered as outliers and are discarded (Volkov & Svergun, 2003). The remaining models are superimposed with the reference, and the entire assembly is remapped onto a densely packed grid. For each knot of the grid, an occupancy factor is assigned equal to the number of the beads belonging to any of the superimposed models in the vicinity of the knot, and the knots with non-zero occupancy form a total spread region. The averaged model corresponding to an inter-connected ensemble of the most populated points is evaluated by filtering the map to yield the volume equal to the average excluded volume of all reconstructions. The following enhancements were made to improve the performance of *DAMAVER* in *ATSAS* 2.1.

(i) The occupancy of a grid knot is now calculated as a total overlap volume of a sphere with the diameter equal to the grid edge centred at this grid point with the beads of the entire assembly that are in the vicinity of this point. This is more accurate than the previously employed computation of occupancy as the number of the beads in the vicinity of the given knot.

(ii) The shift of the centre of mass of the averaged model due to filtering is restricted to a maximum of 20% of the R_g value. This restriction keeps the shape similarity (in particular, maintains the

ratio between the principal inertia axes), between the filtered model and the total spread region.

(iii) The CPU time required by a single *SUPCOMB* run is proportional to the product of the numbers of atoms (beads or DRs) in the two files to be aligned. The computational effort required by *DAMAVER* is proportional to the square of the number of files and may take hours of CPU time. Running of *SUPCOMB* is the most time-consuming part of *DAMAVER*, and optimization of *SUPCOMB* is therefore of major importance. A fast version of *SUPCOMB* was written, which re-maps the input structures onto rough grids so that each model is represented by about a hundred beads. These rough representations are first superimposed, then the corresponding transformations (rotations and shifts) are applied to the full models and the alignment is refined in the vicinity of the position and orientation found by the fast alignment procedure.

By default, *DAMAVER* uses the fast version of *SUPCOMB*, but the original (slower, but sometimes more accurate) version may be invoked using the /s key in the command line. The command string damaver /a runs the entire averaging and filtering in batch mode against all PDB files in the current directory. Fig. 4 demonstrates a comparison between several models of HIV-1 reverse transcriptase (Wang *et al.*, 1994) obtained by a *DAMMIN* run in the 'Keep' mode, and the total spread region and the averaged model generated by the fast version of *DAMAVER* from the entire set of solutions in this run.

It should be noted that, although the average model does preserve the most persistent features of the solution, the scattering computed from this model does not usually fit the experimental data. The present version of *DAMAVER* generates a file DAMSTART.PDB, representing the total spread region where the assignment of the core beads corresponding to the map points with the highest densities occupying in total half of the average excluded volume of all reconstructions are fixed as 'particle' beads. This file can be used as an input to *DAMMIN* and refined to yield the final shape of the particle.

6. Calculation of scattering curves from atomic models

The programs *CRYSOL* for X-rays (Svergun *et al.*, 1995) and *CRYSON* (Svergun *et al.*, 1998) for neutrons are widely used for calculating the scattering profiles from atomic models of macro-molecular structures. These programs compute the scattering from an atomic model of the particle in solution as

$$I(s) = \langle |A(\mathbf{s})|^2 \rangle_\Omega = \langle |A_a(\mathbf{s}) - \rho_s A_s(\mathbf{s}) + \delta\rho_b A_b(\mathbf{s})|^2 \rangle_\Omega, \qquad (11)$$

where $A_a(\mathbf{s})$ is the (X-ray or neutron) scattering amplitude from the particle in vacuum, $A_s(\mathbf{s})$ and $A_b(\mathbf{s})$ 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. Equation (11) takes into account that the density of the bound solvent ρ_b may differ from that of the bulk ρ_s leading to a non-zero contrast of the hydration shell $\delta\rho_b = \rho_b - \rho_s$. Given the atomic coordinates, the programs fit the experimental scattering curve by adjusting the excluded volume of the particle and the contrast of the hydration layer surrounding the particle in solution to minimize the discrepancy in equation (6). In the absence of the experimental data, they can predict the theoretical scattering pattern using default or user-defined parameters. The three terms in equation (11) are computed using a multipole expansion similar to equations (9) and (10) to speed up the calculations. Furthermore, the partial amplitudes of the particle in solution, $A_{lm}(s)$, computed by *CRYSOL/CRYSON* can further be used for rapid computation of scattering from complexes (see §7).
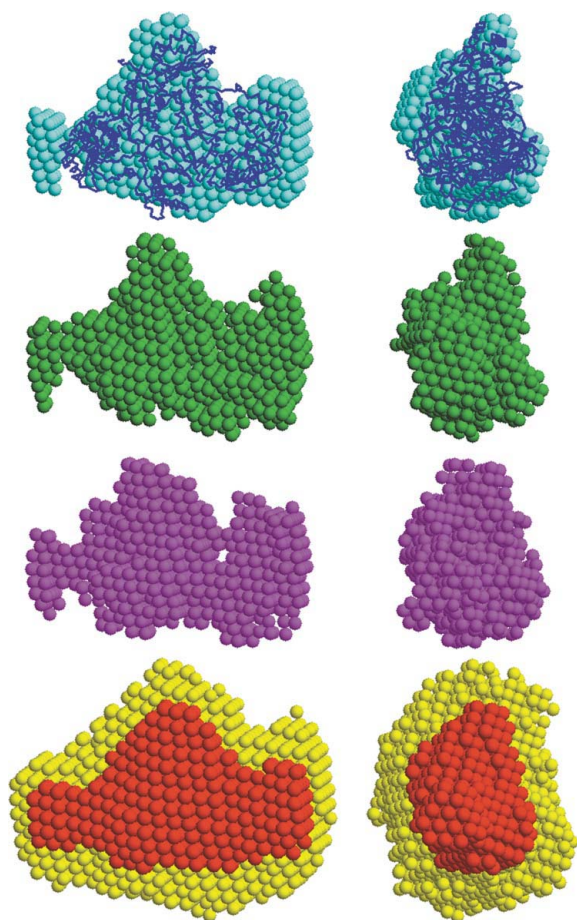

Figure 4
Comparison of *DAMMIN* and *DAMAVER* results (bead models) for HIV-1 reverse transcriptase with the crystallographic model (blue C_α trace). The first-, sixth- and twelfth-ranked models reconstructed by one *DAMMIN* run in the 'Keep' mode are shown in cyan, green and magenta, respectively. The total spread region and the averaged model generated by the *DAMAVER* suite, superimposed, are given in yellow and red. The right-hand view is rotated counterclockwise by 90° around the vertical axis.

computer programs

The functionality of the two programs has recently been further enhanced. The hydrogen atoms are usually missing in atomic models from X-ray crystallography, but their positions can be predicted on all C atoms and on some N atoms and on some O atoms (*i.e.* excluding protonation cases, which are the subject of experimental investigation at very high resolution by X-rays and/or by neutron protein crystallography (Blakeley *et al.*, 2004)]. In the previous versions, H atoms were added based on look-up tables written for amino acid residues, nucleotides and sugars. In the new version (2.6), both programs utilize a dictionary of more than 5000 (bio)chemical components as defined in the PDB (Berman *et al.*, 2000; see http://rcsb-deposi-t.rutgers.edu/het_dictionary.txt). Moreover, the look-up table of heteroatoms was significantly widened. These improvements make it possible to compute solution scattering patterns accurately from virtually any (bio)chemical compound or macromolecule, and not only from proteins, nucleic acids and sugars, as in the previous versions. The programs can also work correctly with the full-atom models containing H atoms or deuterons (*e.g.* with the theoretical models or those solved by NMR or neutron crystallography). Given that *CRYSOL* is now used by different groups to compute the scattering from macromolecules up to relatively wide angles, the calculation of the scattering from the excluded volume $A_s(\mathbf{s})$ was further improved to allow the program to evaluate scattering accurately up to the resolution of about 0.3 nm ($s = 20.0$ nm^{-1}). For *CRYSON*, the possibility of taking into account perdeuteration of the macromolecule was added. Here, individual chains may have a different degree of perdeuteration specified by the user, and the H/D exchange of the perdeuterated chains in solutions with different D$_2$O concentrations is taken into account in a similar way as that of the protonated chains (Svergun *et al.*, 1998).

To facilitate use of the program for screening of multiple models, the possibility of running *CRYSOL* and *CRYSON* in batch mode from the command line was added. The most frequently used parameters are transmitted from a string of keys with key values, similar to *DAMMIN* and *GASBOR*. Wild cards in file names are allowed so that *e.g.* a thousand models from a molecular dynamics trajectory can be tested against the experimental data by a single command.

CRYSOL was recently employed to compile a database of the scattering patterns from about 10 000 proteins and biologically active oligomers with known high-resolution structures in the crystal (Sokolova *et al.*, 2003). This database, DARA, is available at the EMBL Web site (http://dacha.embl-hamburg.de/dara.php) for a fast search of structural neighbours of the given protein based on its experimental SAXS pattern.

7. Rigid-body modelling programs

Macromolecular complexes mediate most of the fundamental biological processes and the focus of modern structural biology shifts towards their study. Often, the structures of individual components are available or can be more easily determined than the high-resolution structures of complexes. Rigid-body modelling against SAXS or SANS data is, together with cryo-electron microscopy based docking (Sali *et al.*, 2003), one of the most promising approaches to construct three-dimensional models of complexes from their components. *ATSAS* 2.1 includes several programs that enable advanced interactive and automated rigid-body modelling of macromolecular complexes.
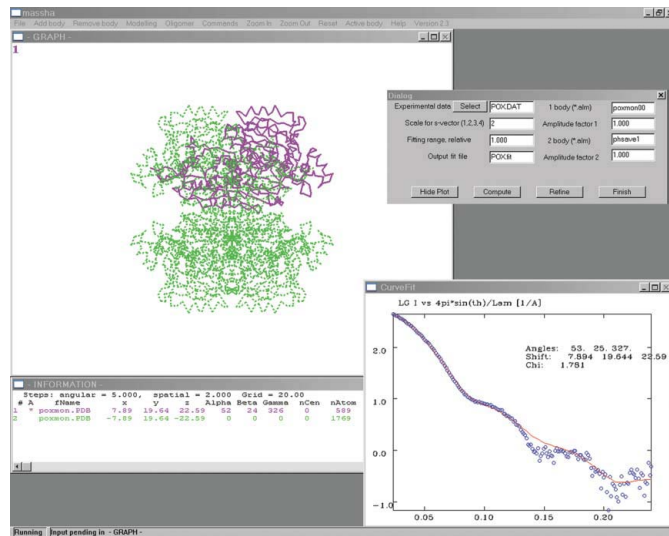

Figure 5
The user interface of *MASSHA*, displaying oligomeric rigid-body modelling (*P*222 symmetry). The refined structure of pyruvate oxidase is displayed as magenta (monomer part) and dashed green C$_\alpha$ chains. The best fit obtained by rigid-body modelling is shown in the bottom right-hand corner.

7.1. Manipulations with three-dimensional atomic models and interactive modelling

The program package *MASSHA* (Konarev *et al.*, 2001) enables the display and manipulation of high-resolution atomic structures and low-resolution models, represented as smooth envelopes or ensembles of beads. This package is not aimed at providing comprehensive rendering similar to macromolecular graphics programs like *RASMOL* (Sayle & Milner-White, 1995) or *VMD* (Humphrey *et al.*, 1996). Instead, *MASSHA* is coupled to computational modules to compute scattering from complex particles rapidly, which opens the possibility for rigid-body refinement of the quaternary structure of macromolecular complexes. The idea of rigid-body modelling is best illustrated by considering a complex of two subunits A and B with known atomic structures. If one fixes subunit A while translating and rotating subunit B, the scattering intensity of the complex is

$$I(s, \alpha, \beta, \gamma, \mathbf{u}) = I_a(s) + I_b(s) + 4\pi^2 \sum_{L=0}^{\infty} \sum_{m=-l}^{l} \mathrm{Re}\,[A_{lm}(s)C_{lm}^*(s)], \quad (12)$$

where $I_a(s)$ and $I_b(s)$ are the scattering intensities from A and B, respectively. The $A_{lm}(s)$ are partial amplitudes of the fixed subunit A, and the $C_{lm}(s)$ are those of subunit B rotated by the Euler angles α, β, γ, and translated by a vector $\mathbf{u}$. The structure and the scattering intensity from such a complex depend on the six positional and rotational parameters and these can be refined to fit the experimental scattering data. The algorithms (Svergun, 1991, 1994) allow one to evaluate the amplitudes $C_{lm}(s)$ rapidly and thus the intensity $I(s, \alpha, \beta, \gamma, \mathbf{u})$ for arbitrary rotations and displacements of the second subunit (the amplitudes from both subunits in reference positions must be pre-computed using *CRYSOL* or *CRYSON*). Equation (12) 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. The computational modules implementing these fast algorithms are coupled with *MASSHA* to allow interactive modelling, where the subunits can be translated and rotated as rigid bodies while observing corresponding changes in the fit to the experimental data. An automated refinement mode is also available for performing an exhaustive search in the vicinity of the current configuration. Previous

versions of *MASSHA* allowed for modelling of hetero- or homo-dimeric complexes (imposing a *P*2 symmetry for the latter case). In the recent version (2.3), additional types of symmetric complexes are allowed, including point symmetry *P*3 to *P*6, and *P*22 to *P*62 (Fig. 5). When a monomer is moved/rotated by the user, its symmetry mates are generated automatically, the scattering from the entire complex is computed and the fit to the experimental data is displayed. For non-symmetric complexes, the possibility of multi-subunit modelling is added, allowing manipulations with up to seven distinct structures. The user may change the position/orientation of an individual subunit or select a group consisting of several subunits. Upon movement or rotation of the subunit/group, the intensity is recalculated and the fit to the experimental data is displayed. Both symmetric and multibody modelling options also allow for automated local refinement by an exhaustive search in the vicinity of the current configuration. Following the criteria introduced by Petoukhov & Svergun (2005), the final model has to be interconnected and display no steric clashes.

7.2. Global modelling using SAXS, SANS and NMR

The program *MASSHA* is convenient for interactive modelling but offers only limited possibilities for automated search. The use of fast computational algorithms based on spherical harmonics (Svergun, 1994) allows one to design efficient global refinement methods. Depending on the complexity of the object, different approaches can be employed for the global search of the optimum configuration of subunits fitting the experimental data (Petoukhov & Svergun, 2005). *ATSAS* 2.1 includes the program *DIMFOM*, which performs an exhaustive grid search for the modelling of hetero- and homodimeric particles whereby one monomer is rolled on the surface of the other. For this, the shapes of the two monomers are represented by *CRYSOL*-generated angular envelope functions $F(\omega)$, where ω is the solid angle in real space. Another program, *GLOBSYMM*, also utilizing a 'brute force' modelling approach, performs quaternary structure analysis of symmetric oligomers formed by identical subunits in terms of position and orientation of one reference monomer on representative spatial and angular grids. The entire oligomer is constructed by appropriate symmetry operations. The program *BUNCH* allows one to study the configuration of multidomain proteins. It employs a combined rigid-body and *ab initio* modelling approach to search for a spatial arrangement of the domains with known high-resolution structure and possible conformations of flexible linkers. The latter are represented as chains composed by dummy residues (Svergun *et al.*, 2001; Petoukhov *et al.*, 2002) connecting the appropriate termini in rigid domains. The program employs SA to move/rotate the domains as rigid bodies by simultaneously allowing for changes of the local conformation of the DR chains representing the linkers.

A comprehensive algorithm for the modelling of macromolecular complexes using high-resolution structures of individual subunits (or domains) is implemented in the program *SASREF*. The program employs an SA protocol (Kirkpatrick *et al.*, 1983; Press *et al.*, 1992; Ingber, 1993) to find the positions and orientations of the subunits forming an interconnected assembly without steric clashes, while minimizing the discrepancy between the calculated and the experimental scattering profiles. The minimization procedure starts from an arbitrary arrangement of subunits, *e.g.* from that in a tentative model of the complex or just from all subunits centred at the origin in their reference orientations. It is possible to fix selected subunits at their starting positions and orientations to preserve known substructures. A single modification of the assembly is done by rotation of a randomly selected subunit by an arbitrary angle $\varphi < \varphi_{\max}$ about a

rotation axis followed by a random shift $r < r_{\max}$ along an arbitrary direction. If the scattering data sets from the partial constructs are also available, the program performs simultaneous fitting of the multiple data sets, which significantly increases the information content and thus the reliability of the modelling. *SASREF* allows one to account for the known interfaces (*e.g.* binding sites) between subunits by restraining the correspondent inter-residue distances. Information about symmetry and/or anisometry of the complex, if available, is taken into account.

The original version of *SASREF* (Petoukhov & Svergun, 2005) was primarily oriented towards SAXS-based modelling. Given the fact that the contrasts of the individual subunits can be effectively varied in neutron scattering using H_2O/D_2O mixtures and/or selective deuteration, and the contrast variation in SANS provides valuable information about complexes (Koch & Stuhrmann, 1979; Zaccai & Jacrot, 1983; Wall *et al.*, 2000), the possibility to account for contrast-variation SANS data was added. Given the scattering amplitudes computed by *CRYSOL* and *CRYSON*, the present version of *SASREF* is able to fit simultaneously X-ray and neutron scattering data, whereas the latter curves can be recorded at different contrasts and also for complexes containing selectively deuterated subunits.

Using rigid-body modelling against SAXS or SANS data, mutual positions of individual subunits can usually be determined rather accurately, but the scattering data may be less sensitive to the orientational parameters. This rotational uncertainty in SAS data is reduced using residual dipolar couplings (RDC) in NMR (Mattinen *et al.*, 2002), which yield long-range order information about the orientation of the subunits. The RDCs can be measured by aniso-tropic tumbling of macromolecules in a liquid-crystalline environment (Guntert, 1998; Prestegard *et al.*, 2000). RDCs in partially aligned molecules in particular give the information on angles between globally defined axes in the molecule, namely those of the magnetic susceptibility tensor (Guntert, 1998; Prestegard *et al.*, 2000). For rigid-body analysis, RDCs provide information about the mutual orientation of subunits in a complex with a fourfold degeneracy corresponding to rotations by 180° about the three orthogonal principal axes, and an option was added to *SASREF* to account for the RDC data, if available. Generation of the initial approximation and all modifications of the model during the SA minimization in *SASREF* keep the principal axes of the relevant subunits compatible with those allowed by RDCs within some degree of tolerance (usually, 5–10°).

The above rigid-body refinement programs allow one to use SAXS and SANS effectively to construct models of complexes from the high-resolution structures of subunits. In particular, the enhanced version of *SASREF* in *ATSAS* 2.1 is a universal program with broad possibilities, like simultaneous fitting of X-ray and contrast-variation neutron scattering data, and accounting for information from other methods in the form of intersubunit contacts and orientation.

8. Conclusions

The package *ATSAS* 2.1 encompasses a set of programs allowing one to perform comprehensive analysis of SAXS and SANS data from isotropic systems. These programs are constantly being modified following feedback from users and the functionality of most of the programs has been significantly enhanced in release 2.1 of *ATSAS*, described in the present paper. In future development, further integration and automation of the programs is planned with the ultimate aim of creating an expert system capable of automated high-throughput data processing and interpretation. The programs

computer programs

included in *ATSAS* are publicly available from the EMBL Web site for academic users: http://www.embl-hamburg.de/ExternalInfo/Research/Sax/software.html.

References

Baxter, R. J. (1968). *J. Chem. Phys.* **49**, 2270–2273.
Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.
Bevington, P. B. (1969). *Data Reduction and Error Analysis for the Physical Sciences*. New York: McGraw-Hill.
Blakeley, M. P., Cianci, M., Helliwell, J. R. & Rizkallah, P. J. (2004). *Chem. Soc. Rev.* **33**, 548–557.
Boulin, C., Kempf, R., Koch, M. H. J. & McLaughlin, S. M. (1986). *Nucl. Instrum. Methods A*, **249**, 399–407.
Brandstetter, H., Kim, J.-S., Groll, M. & Huber, R. (2001). *Nature (London)*, **414**, 466–470.
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.
Dewhurst, C. (2002). *GRASP* software package. Institute Laue–Langevin, Grenoble, France.
Feigin, L. A. & Svergun, D. I. (1987). *Structure Analysis by Small-Angle X-ray and Neutron Scattering*. New York: Plenum Press.
Ghosh, R. E. (1989). Institute Laue–Langevin Internal Publication 89GH02T, Grenoble, France.
Glatter, O. (1977). *J. Appl. Cryst.* **10**, 415–421.
Golub, G. H. & Reinsch, C. (1970). *Numer. Math.* **14**, 403–420.
Guinier, A. (1939). *Ann. Phys. (Paris)*, **12**, 161–237.
Guntert, P. (1998). *Q. Rev. Biophys.* **31**, 145–237.
Hammersley, A. P. (1995). ESRF Internal Report Exp/AH/95-01, Grenoble, France.
Heenan, R. K. (1999). *FISH*, program for peak analysis. Rutherford Appleton Laboratory Internal Publication 89-129, Didcot, UK.
Heenan, R. K., Penfold, J. & King, S. M. (1997). *J. Appl. Cryst.* **30**, 1140–1147.
Heller, W. T., Abusamhadneh, E., Finley, N., Rosevear, P. R. & Trewhella, J. (2002). *Biochemistry*, **41**, 15654–15663.
Hinde, A. L. (2004). *J. Appl. Cryst.* **37**, 1020–1024.
Hiragi, Y., Sano, Y. & Matsumoto, T. (2003). *J. Synchrotron Rad.* **10**, 193–196.
Homan, E., Konijnenburg, M., Ferrero, C., Ghosh, R. E., Dolbnya, I. P. & Bras, W. (2001). *J. Appl. Cryst.* **34**, 519–522.
Humphrey, W. F., Dalke, A. & Schulten, K. (1996). *J. Mol. Graph.* **14**, 33–38.
Ingber, L. (1993). *Math. Comput. Model.* **18**, 29–57.
Keiderling, U. (1997). *Physica B*, **234–236**, 1111–1113.
King, S. M. & Heenan, R. K. (1996). Rutherford Appleton Laboratory Internal Publication RAL-TR-96-036, Didcot, UK.
Kirkpatrick, S., Gelatt, C. D. Jr & Vecci, M. P. (1983). *Science*, **220**, 671–680.
Koch, M. H. J. & Stuhrmann, H. B. (1979). *Methods Enzymol.* **59**, 670–706.
Konarev, P. V., Petoukhov, M. V. & Svergun, D. I. (2001). *J. Appl. Cryst.* **34**, 527–532.
Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. (2003). *J. Appl. Cryst.* **36**, 1277–1282.

Kozin, M. B. & Svergun, D. I. (2001). *J. Appl. Cryst.* **34**, 33–41.
Larson, H. J. (1975). *Statistics: an Introduction*. New York: John Wiley.
Lawson, C. L. & Hanson, R. J. (1974). *Solving Least Squares Problems*. Englewood Cliffs, NJ: Prentice-Hall.
Malfois, M. & Svergun, D. I. (2000). *J. Appl. Cryst.* **33**, 812–816.
Mattinen, M. L., Paakkonen, K., Ikonen, T., Craven, J., Drakenberg, T., Serimaa, R., Waltho, J. & Annila, A. (2002). *Biophys J.* **83**, 1177–1183.
Pedersen, J. S. (2002). *Neutrons, X-rays and Light*, edited by P. Lindner & T. Zemb, pp. 127–144. Amsterdam: North Holland.
Petoukhov, M. V., Eady, N. A., Brown, K. A. & Svergun, D. I. (2002). *Biophys J.* **83**, 3113–3125.
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.
Porod, G. (1982). *Small-Angle X-ray Scattering*, edited by O. Glatter & O. Kratky, pp. 17–51. London: Academic Press.
Press, W. H., Teukolsky, S. A., Wetterling, W. T. & Flannery, B. P. (1992). *Numerical Recipes*. Cambridge University Press.
Prestegard, J. H., al-Hashimi, H. M. & Tolman, J. R. (2000). *Q. Rev. Biophys.* **33**, 371–424.
Sali, A., Glaeser, R., Earnest, T. & Baumeister, W. (2003). *Nature (London)*, **422**, 216–225.
Sayle, R. A. & Milner-White, E. J. (1995). *Trends Biochem. Sci.* **20**, 374–376.
Semenyuk, A. V. & Svergun, D. I. (1991). *J. Appl. Cryst.* **24**, 537–540.
Shannon, C. E. & Weaver, W. (1949). *The Mathematical Theory of Communication*. Urbana: University of Illinois Press.
Sokolova, A. V., Volkov, V. V. & Svergun, D. I. (2003). *J. Appl. Cryst.* **36**, 865–868.
Strunz, P., Saroun, J., Keiderling, U., Wiedenmann, A. & Przenioslo, R. (2000). *J. Appl. Cryst.* **33**, 829–833.
Stuhrmann, H. B. (1970). *Z. Physik. Chem. Neue Folge*, **72**, 177–198.
Svergun, D. I. (1991). *J. Appl. Cryst.* **24**, 485–492.
Svergun, D. I. (1992). *J. Appl. Cryst.* **25**, 495–503.
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. & Koch, M. H. J. (2003). *Rep. Prog. Phys.* **66**, 1735–1782.
Svergun, D. I., Konarev, P. V., Volkov, V. V., Koch, M. H. J., Sager, W. F. C., Smeets, J. & Blokhuis, E. M. (2000). *J. Chem. Phys.* **113**, 1651–1665.
Svergun, D. I., Petoukhov, M. V. & Koch, M. H. J. (2001). *Biophys. J.* **80**, 2946–2953.
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., Semenyuk, A. V. & Feigin, L. A. (1988). *Acta Cryst.* A**44**, 244–250.
Svergun, D. I., Volkov, V. V., Kozin, M. B. & Stuhrmann, H. B. (1996). *Acta Cryst.* A**52**, 419–426.
Vainshtein, B. K. (1966). *Diffraction of X-rays by Chain Molecules*. Amsterdam: Elsevier.
Vigil, D., Gallagher, S. C., Trewhella, J. & Garcia, A. E. (2001). *Biophys J.* **80**, 2082–2092.
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, J., Smerdon, S. J., Jager, J., Kohlstaedt, L. A., Rice, P. A., Friedman, J. M. & Steitz, T. A. (1994). *Proc. Natl Acad. Sci. USA*, **91**, 7242–7246.
Winn, M. D., Ashton, A. W., Briggs, P. J., Ballard, C. C. & Patel, P. (2002). *Acta Cryst.* D**58**, 1929–1936.
Zaccai, G. & Jacrot, B. (1983). *Annu Rev. Biophys. Bioeng.* **12**, 139–157.