

ATSAS 2.1 – towards automated and web-supported small-angle scattering data analysis

Maxim V. Petoukhov,^{a,b} Peter V. Konarev,^{a,b} Alexey G. Kikhney^a and Dmitri I. Svergun^{a,b*}

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

Small-angle scattering (SAS) is frequently employed for screening large numbers of samples and for studying these samples under different conditions, including space- and time-resolved analysis. These measurements produce immense amounts of data, especially on modern high-flux and high-brilliance sources (*e.g.* third-generation synchrotrons). In biological SAS, like high-throughput macromolecular crystallography, large-scale analysis of proteins and macromolecular complexes is also emerging. Automation of data analysis becomes an indispensable prerequisite for adequate evaluation of high-throughput SAS experiments. Here a prototype of an automated data-analysis system for isotropic solution scattering based on the further development of the programs belonging to the package *ATSAS 2.1* is reported. This system allows the major analysis tasks starting from the raw data processing and, for monodisperse systems, finishing with a three-dimensional model, to be performed automatically. Convenient web interfaces for the online use of individual *ATSAS* programs are also provided.

© 2007 International Union of Crystallography
Printed in Singapore – all rights reserved

1. Introduction

Small-angle scattering (SAS) of X-rays and neutrons (SAXS and SANS) are methods which often provide significant amounts of experimental data. This is related to the fact that SAS is frequently employed for screening large numbers of samples, and also to the use of SAS for the analysis of kinetic processes and structural responses to changes in external conditions (temperature, pressure, chemical modifications *etc.*). During the last decade, the amount of data generated by SAS increased dramatically thanks to the use of high-brilliance synchrotron sources. This applies *e.g.* to the use of micro-focus synchrotron X-ray scattering for space-resolved measurements and also to the studies of macromolecular solutions. Admittedly, SAXS still lags behind other techniques such as high-throughput macromolecular crystallography in terms of automation, but the tendency towards high-throughput SAXS is clearly seen. One may quote for example a 'Lab-on-a-Chip' approach being developed in Copenhagen, Denmark (<http://www.dfuni.dk/index.php/SAXS-in-a-microTAS/2711/0/>) or medium-scale screening of membrane proteins from *Thermotoga maritima* for solubility in different detergents (Columbus *et al.*, 2006).

The challenge of the rapidly growing amount of experimental data requires adequate means for data evaluation. Automated data analysis and interpretation tools would already have increased the efficiency of structural SAXS/SANS studies, and such tools are expected to become indispensable in the near future. Several publicly available program packages have been developed (mostly at large-scale facilities) to analyse SAXS/SANS data. They include various data reduction and processing tools, both general and oriented to specific objects (Keiderling, 1997; Homan *et al.*, 2001; Dewhurst, 2002; Hiragi *et al.*, 2003; Davies, 2006). Modelling and interpretation

programs also exist for different types of systems (*e.g.* Hammersley, 1995; Chacon *et al.*, 1998; Heenan, 1999; Walther *et al.*, 2000). In most cases, substantial user intervention and changing between programs is required during the data analysis/interpretation process. In the present paper we report a prototype of an automated data analysis system for isotropic solution scattering based on the programs developed in the package *ATSAS 2.1* (Konarev *et al.*, 2006). This system allows the user to accomplish the data analysis tasks starting from the raw data processing and finishing with a low-resolution three-dimensional (3D) structural model. Moreover, convenient web interfaces for the use of individual *ATSAS* programs are provided.

ATSAS 2.1 is a computer package primarily oriented to the analysis of solutions of biological macromolecules, but it can also be used for non-biological systems yielding one-dimensional (1D) isotropic scattering patterns. *ATSAS 2.1* includes the Windows-based data processing and reduction program *PRIMUS* (Konarev *et al.*, 2003), which computes overall structural parameters and provides interfaces to programs for data manipulation, component and peak analysis, and modelling by simple geometrical bodies. Enhanced 3D modelling can be done using *DAMMIN* (Svergun, 1999) and *MONSA* (Svergun & Nierhaus, 2000) for low-resolution shape determination of single and multi-component particles, respectively, or *GASBOR* (Svergun *et al.*, 2001) for *ab initio* domain structure determination of proteins from X-ray data by representing them as ensembles of identical dummy residues. The programs *CRYSOL* for X-rays (Svergun *et al.*, 1995) and *CRYSON* for neutrons (Svergun *et al.*, 1998) allow the user to calculate the scattering profiles from atomic models of macromolecular structures. A rigid body modelling suite including programs *MASSHA* and *SASREF* is also available to characterize macromolecular complexes in terms of the structure of subunits (Konarev *et al.*, 2001; Petoukhov & Svergun, 2005). Most of

the programs belonging to *ATSAS* run on multiple hardware platforms (Windows PC, Linux, Mac OSX, different UNIX flavours). The package and its components are publicly available for academic users from the EMBL website (<http://www.embl-hamburg.de/ExternalInfo/Research/Sax/software.html>).

As seen from the description above, *ATSAS* 2.1 provides useful tools covering the major analysis steps of the experimental isotropic scattering data. At present, individual programs are invoked by the user interactively. For Windows versions this has to be done from the corresponding graphical user interface (GUI); for the Unix-based versions the programs are run from the command line. We have developed a Windows-based prototype of an automated pipeline for high-throughput solution scattering data analysis establishing compatibility and interactions between the individual *ATSAS* programs. New modules have been written for automated analysis such that the system is able to run largely in parallel to the data collection without user intervention. In addition, web interfaces have been created to run major *ATSAS* programs and test online access to these programs is available at the EMBL Hamburg website. In the present paper, brief reminders are given of the functionality of the existing *ATSAS* programs (the readers are referred to the original papers for detail), and the new automated modules and the web interfaces are described in more detail.

2. Automated data reduction

Automated data reduction and normalization is an indispensable first step for any high-throughput analysis system. For isotropic scattering this step involves radial averaging of the scattering data recorded on a two-dimensional (2D) detector and normalization against appropriate monitor values. The data reduction procedure employed at the X33 beamline is comprehensively described elsewhere (Konarev *et al.*, 2006). Briefly, a raw image data file (in this case, from a MAR345 image plate detector) is transformed into 1D arrays of scattering intensities $I(s)$ and their associated errors as a function of the modulus of the scattering vector [$s = (4\pi/\lambda)\sin\theta$, where λ is the wavelength and 2θ the scattering angle]. The transformation is performed by integration over concentric rings with respect to the pre-defined beam centre position. The beam centre and the angular axis are determined from the scattering profile of silver behenate using the program *FIT2D* (Hammersley, 1995). Finally the data are normalized to the transmitted sample intensity and to the collection time. Previously, this step was done using a pop-up window of *PRIMUS* and this procedure had to be run for each new measurement or group of measurements (Konarev *et al.*, 2003). The reduction procedure is now automated by the program *AUTOMAR*, which only requires an initialization file in which the required parameters and working directories are specified. *AUTOMAR* runs in the background and permanently scans the raw data directory. When it finds the new raw image data file(s), they are read and transformed. The reduced files in ASCII format are stored in the processed data directory for subsequent analysis.

The next data processing step involves subtraction of the background. In solution scattering, the background is defined by the scattering from the pure solvent, and, especially for biological samples, the useful difference may be rather small compared to the solvent scattering. Further, on synchrotrons, minor movements of the incident beam during the experiment may lead to instabilities of the background subtraction. Typically, the measurement of the solute is therefore surrounded by two solvent (buffer) measurements. Moreover, two or more measurements of the sample are sometimes carried

out in order to monitor possible radiation damage. In the program *PRIMUS*, the sample or buffer averaging and sample – buffer subtraction was done interactively, by loading the experimental files and using the ‘operation’ buttons of the dialog toolbox (Konarev *et al.*, 2003). Possible instabilities in the background and monitoring radiation damage were done by visual inspection of the scattering patterns. Subtraction of the averaged buffer works well in most cases but this still requires user intervention, which is unacceptable in high-throughput mode. A program *AUTOSUB* has been developed for automated background subtraction, and operates on the files processed by *AUTOMAR*. The program recognizes sample and background measurements from the headers of the reduced files, and analyses the subtracted data using the backgrounds (‘buffers’ for solution scattering) measured before and after the sample. First, a statistical analysis is done to characterize the stability of the background by comparing the two buffers using a standard F-test (Bevington, 1969). Similarly, the F-test is done on successive sample measurements to monitor possible radiation damage, if relevant. If the compared files are statistically indistinguishable, appropriate averaging operations are done and the averaged background is subtracted. If not, three possible subtractions are considered: (i) sample – buffer 1; (ii) sample – buffer 2; (iii) sample – $\frac{1}{2}(\text{buffer 1} + \text{buffer 2})$, and for each of them numerical goodness criteria are computed. These criteria include requirements of the absence of systematically negative portions in the subtracted curve and proximity of the sample and background scattering at higher angles where the useful signal is expected to be relatively small. Further, the radii of gyration are calculated by the program *AUTORG*, which also yields a quality estimate of the Guinier fit (see §3). Finally, the total estimate is composed for each subtracted curve from these criteria to select the best subtracted curve. The *AUTOSUB* procedure combined with the program *AUTORG* (see §3) runs without user intervention.

3. Automated radius of gyration calculation from solution scattering data

The radius of gyration (R_g), a classical parameter obtained from the scattering data, is computed using the well known Guinier approximation (Guinier, 1939)

$$I(s) = I(0) \exp(-s^2 R_g^2/3). \quad (1)$$

The value of R_g is estimated from the linear fit of $\ln[I(s)]$ versus s^2 (a Guinier plot), which is valid for sufficiently small scattering vectors (in the range up to $sR_g \lesssim 1.3$). The R_g value provides an estimate of the overall size of particles, which is important for further automated data handling. Moreover, linearity of the Guinier plot is a sensitive indicator of the quality of the experimental data, and deviations from linearity usually point to strong interference effects, polydispersity of the samples or improper background subtraction. Despite the simplicity of the Guinier formula, automated computation of R_g is not a trivial task, in particular because of uncertainty in the fitting interval. Visual inspection is most often used to select the range of the Guinier fit, and this interactive fitting can be conveniently done in several packages, including *PRIMUS*. However, despite the importance of R_g for SAXS/SANS, publicly available programs for automated R_g determination do not seem to be available. We have developed a program *AUTORG* for a fully automated determination of R_g from the scattering data. The program also estimates the quality of the fit and provides information for other modules of the automated pipeline. The current version of the program is designed for

Windows OS and can be run in two ways: as a menu-driven application with a simple GUI or as a console application with command line input (the latter is incorporated in the automated system).

AUTORG works with the experimental data files in standard ASCII format. First, the program selects the data range suitable for the Guinier approximation. For this, the initial portion of the data is analysed and any range showing unreasonable upwards or downwards trends (e.g. caused by the beam stop or strong background near the primary beam) is discarded. Then the data range where the scattering intensity decays by an order of magnitude is taken. A cubic parabola is drawn in this range using a log scale of intensity to analyse the curvature and possible inflection points suggesting 'non-monodisperse' behaviour, and the range is refined when necessary. Then a search of all possible intervals for Guinier plots starts in the selected range: for each interval (longer than a given minimum interval length in points, usually, between 5 and 15) a weighted linear fit is calculated by least squares and R_g is computed. For each interval (s_{\min} , s_{\max}), the conditions $s_{\min}R_g < 1$ and $s_{\max}R_g < 1.3$ are checked and the absence of systematic variations is verified, in which case the interval is considered consistent. If no consistent intervals are found, the program tries to find intervals with weakened sR_g conditions, but simultaneously reduces the estimate of the data quality.

Each consistent interval is rated according to its length (number of points fitted) and discrepancy (root-mean-square deviation of the fit), and the interval with the best rating is selected. The accuracy of R_g is estimated by taking into account not only the error propagation in the selected fit as usual but also by accounting for the deviation of R_g values calculated from other consistent intervals, accounting to some extent for systematic errors in the R_g determination. An estimate of the overall data quality is then expressed by taking into account several criteria: (i) how many consistent intervals were found; (ii), whether the sR_g conditions were weakened or not; (iii) how many starting points were discarded; (iv) whether there is an indication of effects like aggregation; (v) how accurate is the value of R_g . This

estimate is then made available to other programs in the pipeline, in particular to *AUTOSUB* for selecting the optimum subtraction of the background. *AUTORG* tries to translate the perceptual criteria used during interactive R_g analysis by Guinier approximation into an algorithm to compute R_g and to estimate the quality of the fit. The program has several tunable parameters, such as the intensity decay in the fitting range, the minimum interval length in points, the worst acceptable $s_{\min}R_g$ and $s_{\max}R_g$ limits, and the length and discrepancy weights used for the interval rating. These parameters are currently tuned to provide the most stable results, but in future releases can be adjusted by the user. The console version of *AUTORG* using default parameters was tested on numerous data sets and the results were compared with those of manual R_g determination with *PRIMUS*; in the vast majority of cases the automated system yielded the same results as those obtained interactively by an experienced user. Currently the automated mode covers cases of monodisperse or moderately polydisperse systems with sufficiently high contrast, but further work is planned to extend its range of applicability.

4. Evaluation of shape and overall parameters

For monodisperse systems of particles further integral parameters and the particle shape can be automatically computed. After the R_g value is determined and the intensity is extrapolated to zero angle, the excluded volume of the particle, V , can be computed using the Porod equation (Porod, 1982),

$$V = \frac{2\pi^2 I(0)}{Q}, \quad Q = \int_0^\infty [I(s) - K_4] s^2 ds, \quad (2)$$

where Q is the Porod invariant and K_4 is a constant determined to ensure the asymptotical intensity decay proportional to s^{-4} at higher angles. The program *AUTOPOROD* uses, if possible, the portion of the intensity which decays by about two orders of magnitude compared to $I(0)$ for the calculation of K_4 and Q , whereas the truncation effect [integration up to a finite upper limit of s in equation (2)] is taken into account as described by Rolbin *et al.* (1973).

For a simplified but fast estimate of the particle shape, a three-parameter fit using the program *BODIES* is employed. The program finds the best fits from simple geometrical bodies (three-axial ellipsoids, ellipsoids of revolution, cylinders, hollow spheres, hollow cylinders, elliptic cylinders and rectangular prisms) to the experimental data. The calculated values of R_g and V are used to generate the initial approximation for the fitting and the best parameters of the bodies are determined by a non-linear minimization procedure. To automatically calculate the distance distribution function of the particle and determine its maximum size, an automated version of the program *GNOM* (Svergun, 1992) was developed. In the original version of *GNOM* the maximum particle size D_{\max} is a user-defined parameter and successive calculations with different D_{\max} values are required to select its optimum value. This optimum D_{\max} should provide a smooth real-space distance distribution function $p(r)$ such that $p(D_{\max})$ and its first derivative $p'(D_{\max})$ are approaching zero, and the back-transformed intensity from the $p(r)$ fits the experimental data. In the program *AUTOGNOM*, multiple *GNOM* runs are performed to find the optimum D_{\max} and $p(r)$ function. The D_{\max} values ranging from $2R_g$ to $4R_g$ are scanned with a step of $0.1R_g$, where R_g is the radius of gyration provided by *AUTORG*. The calculated $p(r)$ functions for different D_{\max} and corresponding fits to the experimental curves are compared using the perceptual criteria of *GNOM* (Svergun, 1992), where the smoothness of $p(r)$, absence of systematic deviations in the fit and other quantities characterizing the

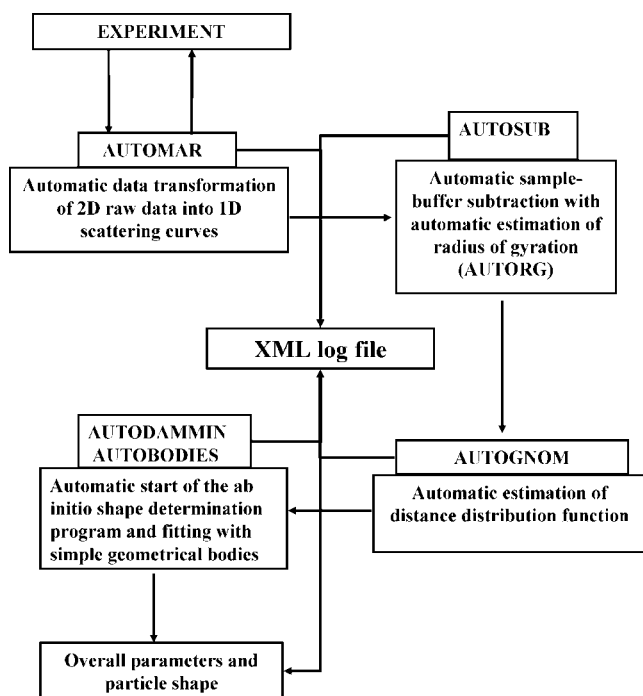


Figure 1

A prototype of the automated data processing system.

conference papers

solution are merged into a total quality estimate. Moreover, the appropriately normalized value of $p'(D_{\max})$ is added to the estimate to ensure that the $p(r)$ function goes smoothly to zero. The best solution according to *AUTOGNOM* is selected and the function $p(r)$ together with calculated overall parameters is stored. Test computations with *AUTOGNOM* on various systems demonstrated that the program is able to reliably select the maximum size and calculate the $p(r)$ function, with results compatible with those of interactive analysis.

Optionally, the final output file from *AUTOGNOM* is submitted for automated *ab initio* shape determination by the program *DAMMIN* (see §5). However, this step can be omitted, as even in the fast mode *DAMMIN* currently requires about 20–30 min on an average PC, which would create a queue if used for all data files in the high-throughput mode. Work is now under way on creating faster versions of the shape determination programs, in particular, by parallelizing their code.

A scheme of the current prototype of the automated analysis system is presented in Fig. 1. An important part of this system is the storage of the retrieved information and of the history of the data analysis. Besides storing the output information from individual modules in their individual log files in ASCII text format, an XML (Extensible Markup Language, see <http://www.w3.org/xml>) file is also generated which contains the main parameters of the entire data processing and primary analysis cycle. XML is one of the new technologies that provide a solution for sharing information across different computing platforms and presents a practical approach to data categorization and communication. The XML tags are user-defined and this results in fast and convenient browsing. The ability to create your own tagging structure gives the language the possibility to categorize and structure data for both ease of retrieval and ease of display. XML is already being used for publishing as well as for data storage and retrieval, data interchange between heterogeneous platforms, data transformations, and data displays. The log data can be presented in various ways (brief, detailed, grouped specifically) which helps in tracking the data processing. An easy way to publish the log files on the web is also provided.

5. A test web interface for SAXS data analysis and model building

Besides the fully automated mode of operation, an 'expert' mode of the integrated analysis system is also foreseen, allowing experienced users to select the appropriate data analysis strategy and to launch the relevant computational modules separately. As a primary step towards an online 3D model building and validation service, web

Welcome to **DAMMIN** Web-service

GNOM output file

Particle symmetry

DAMMIN mode: FAST SLOW KEEP

Figure 2
A view of a web browser window showing an online *DAMMIN* interface.

interfaces to the most frequently used *ab initio* and rigid body modelling algorithms from the *ATSAS* program package have been created.

5.1. *Ab initio* shape determination

Construction of a low-resolution model of the particle shape *ab initio* is probably the most convenient way of interpreting SAS data from monodisperse solutions. *ATSAS* 2.1 contains an *ab initio* program *DAMMIN* (Svergun, 1999) which represents the particle as a collection of several thousands of densely packed beads and employs simulated annealing (SA) to search for a compact model that fits the low-resolution portion of the data (usually to about 2 nm resolution). A recently added option to run *DAMMIN* in a batch mode from the command line (Konarev *et al.*, 2006) allowed us to implement a simple web interface for launching the program re-

Welcome to **CRY SOL** Web-service

PDB file name Keep all the results

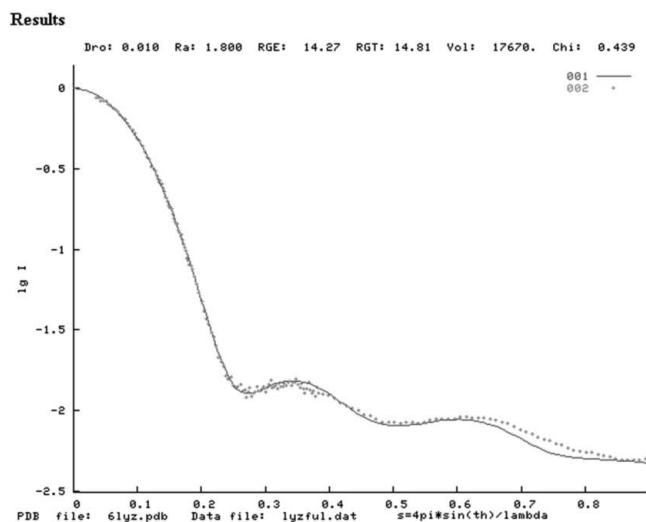
Max order of harmonics Order of Fibonacci grid

Max s value Number of points

Solvent density Contrast of hydration shell

Fit the experimental data

Select angular units
 $4\pi\sin(\theta)/\lambda, \text{\AA}^{-1}$ $4\pi\sin(\theta)/\lambda, \text{nm}^{-1}$ $2\sin(\theta)/\lambda, \text{\AA}^{-1}$ $2\sin(\theta)/\lambda, \text{nm}^{-1}$



The results are available [here](#) as a zip file.
Thank you for using CRY SOL.

Figure 3
Input interface of *CRY SOL* and the output showing the fit to the experimental data.

motely. The user needs to upload the output file from *GNOM*, to specify the expected particle symmetry (*P1* is assumed by default) and to select one of the three *DAMMIN* modes: FAST, SLOW or KEEP [see Konarev *et al.* (2006) for details]. A screenshot of the online *DAMMIN* submission page is presented in Fig. 2. For all the programs running online, current progress is displayed on the screen in text (log file outputs) and graphical (model fits) forms, and the final result is available for download as a single zipped archive file.

5.2. Domain structure analysis of proteins

Another *ab initio* algorithm developed for domain structure determination of proteins from SAXS data represents a protein as a collection of dummy residues (DRs) (Svergun *et al.*, 2001). It takes into account higher resolution data and allows one to get more detailed structural information compared to bead modelling. The DR modelling program *GASBOR* uses SA for fitting either the experimental SAXS data in reciprocal space (Svergun *et al.*, 2001) or the corresponding distance distribution function $p(r)$ provided by *GNOM* in real space (Petoukhov & Svergun, 2003). Like *DAMMIN*, the program has a command line (batch) mode of operation (Konarev *et al.*, 2006) which also facilitates remote job submission. The *GASBOR* web interface is similar to that of *DAMMIN* but has an additional input parameter, the number of DRs in the asymmetric part. The user can also choose between the reciprocal and real space versions to run instead of the three possible modes in *DAMMIN*.

5.3. Solution scattering prediction from high-resolution models

If the high-resolution structure of a macromolecule is available, *ATSAS* provides tools to validate the structural similarity in a crystal and in solution. The programs *CRY SOL* for X-rays (Svergun *et al.*, 1995) and *CRY SON* (Svergun *et al.*, 1998) for neutrons calculate the scattering profiles of macromolecular structures. The programs either fit the experimental scattering curve by adjusting the excluded

volume and the contrast of the hydration layer or predict theoretical scattering patterns using default or user-defined parameters. As *CRY SOL* can be run from the command line, where the most frequently used parameters are transmitted from a string of keys with key values, the online web interface provides nearly full control of the program operation. Typical views of the *CRY SOL* input and output interfaces for experimental data fitting are given in Fig. 3.

5.4. Rigid body modelling

The synergistic use of low-resolution methods like SAS with high-resolution techniques like crystallography or NMR is one of the most promising directions in modern structure studies of macromolecular complexes. In many cases, atomic models of individual components of biological complexes are available, whereas the structure of the entire complex is difficult to analyse with high-resolution methods. Large macromolecular complexes playing key roles in cellular functions are among the most challenging objects for structural studies. Rigid body modelling against SAXS or SANS data is one of the possible options for constructing 3D models of complexes from their components. The most general rigid body modelling algorithm in the *ATSAS* package is implemented in the automated global refinement program *SASREF* (Petoukhov & Svergun, 2005). Here, SA is employed to find an interconnected assembly of subunits with the desired intersubunit interfaces (if known from other methods) but without steric clashes which fits the experimental scattering data. Given the variety of options in *SASREF* (multiple data fitting, symmetry, contacts, possibly different subunits, perdeuterations in contrast variation series *etc.*) the input to the console version of the program is rather complicated. As illustrated in Fig. 4, web browsers also allow one to create convenient interfaces for complex inputs. On the first page (Fig. 4) the user is asked for the total number of the scattering curves, the number of subunits in the asymmetric part and the overall symmetry of the complex. On the next (main) page, general information is provided for each curve and each subunit (Fig. 4). In addition, optional text files containing information on intersubunit contacts and on experimental parameters for the smearing of calculated SANS curves can be uploaded. Finally, the user fills the subunit-curve cross-table specifying the contributions of the subunits to each set of scattering data (presence/absence/perdeuteration) (Fig. 4) and submits the job for remote computation. The user does not need to pre-compute partial scattering amplitudes (required for the off-line version) as the web version of *SASREF* launches *CRY SOL* and *CRY SON* automatically to generate the necessary amplitude files. Currently the contact conditions file has to be provided in text form, but we are planning to add a service for its menu-driven generation.

Web interface for SASREF --- Page 1

Web interface for SASREF --- Page 3 of 3

Perdeuterations of the subunits in each construct (specify "1.0" if the subunit does not appear in the construct)

	sub1.pdb	sub2.pdb	sub3.pdb
xcomp.dat	0.00	0.00	0.00
x-12.dat	0.00	0.00	-1.0
x-23.dat	-1.0	0.00	0.00
nc_0.dat	0.00	0.00	0.00
nc_p50_0.dat	0.00	0.5	0.00
nc_p50_100.dat	0.00	0.5	0.00

Web interface for SASREF --- Page 2 of 3

no. of subunits: 3 - no. of curves: 6 - overall symmetry: P1

Curve	File	D2O fraction	Symmetry	Angular units $\frac{4\pi\sin(\theta)/\lambda}$	Fraction to fit	Setting	Weight	Use a constant?
1	C:\project1\comp.dat	-1.00	P1	A-1	1.00	0	1.00	No
2	C:\project1\x-12.dat	-1.00	P1	A-1	1.00	0	1.00	No
3	C:\project1\x-23.dat	-1.00	P1	A-1	1.00	0	1.00	No
4	C:\project1\nc_0.dat	0	P1	A-1	1.00	1	1.00	Yes
5	C:\project1\nc_p50_0.dat	0	P1	A-1	1.00	1	1.00	Yes
6	C:\project1\nc_p50_100	1.00	P1	A-1	1.00	2	1.00	Yes

Subunit	File	Shift?	Fix?	Symmetry
1	C:\project1\sub1.pdb	Yes	No	P1
2	C:\project1\sub2.pdb	Yes	Yes	P1
3	C:\project1\sub3.pdb	Yes	No	P1

Figure 4
The web interface of *SASREF*. Three web pages are displayed showing successive menus (page 1: overall parameters; page 2: information about scattering patterns and subunits; page 3: cross-table of contrasts).

6. Conclusions

Our long-term objective is to create a high-throughput integrated system for rapid structural analysis of isotropic monodisperse systems covering all the analysis steps from data reduction to automated modelling. Web access to the system would facilitate its use by a broader community by running it remotely or with minimal help from SAXS beamline personnel. The system will employ standard data formats, databases of scattering patterns and modern analysis algorithms. A decision-making block will be designed to select proper analysis actions and to compare concurrent models or suggest experiments reducing the ambiguity of the current model. This system will be primarily oriented towards the analysis of biological macromolecules, but could also be used for non-biological isotropic and partially oriented objects (inorganic, colloidal solutions, poly-

mers in solution and bulk). It is also planned to include automated data analysis of neutron scattering data so that SANS applications will be covered. The present paper describes the first steps towards such an integrated system. The programs will be posted on the EMBL website (<http://dacha.embl-hamburg.de/atsas>) and the system will be further developed based on feedback from the user community.

The authors acknowledge financial support from the EU Framework 6 Programme (Design Study SAXIER, RIDS 011934).

References

- Bevington, P. B. (1969). *Data reduction and error analysis for the physical sciences*. New York: McGraw-Hill.
- Chacon, P., Moran, F., Diaz, J. F., Pantos, E. & Andreu, J. M. (1998). *Biophys. J.* **74**(6), 2760–2775.
- Columbus, L., Lipfert, J., Klock, H., Millett, I., Doniach, S. & Lesley, S. A. (2006). *Protein Sci.* **15**, 961–975.
- Davies, R. J. (2006). *J. Appl. Cryst.* **39**, 267–272.
- Dewhurst, C. (2002). *GRASP* software package. Institute Laue–Langevin, Grenoble, France.
- Guinier, A. (1939). *Ann. Phys. (Paris)*, **12**, 161–237.
- 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.
- 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.
- Keiderling, U. (1997). *Physica B*, **234–236**, 1111–1113.
- Konarev, P. V., Petoukhov, M. V. & Svergun, D. I. (2001). *J. Appl. Cryst.* **34**, 527–532.
- Konarev, P. V., Petoukhov, M. V., Volkov, V. V. & Svergun, D. I. (2006). *J. Appl. Cryst.* **39**, 277–286.
- Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. (2003). *J. Appl. Cryst.* **36**, 1277–1282.
- Petoukhov, M. V. & Svergun, D. I. (2003). *J. Appl. Cryst.* **36**, 540–544.
- Petoukhov, M. V. & Svergun, D. I. (2005). *Biophys. J.* **89**(2), 1237–1250.
- Porod, G. (1982). *General theory. Small-angle X-ray scattering*, edited by O. Glatter and O. Kratky, pp. 17–51. London: Academic Press.
- Rolbin, Y. A., Kayushina, R. L., Feigin, L. A. & Schedrin, B. M. (1973). *Kristallografiya*, **18**, 701–705. (In Russian.)
- Svergun, D. I. (1992). *J. Appl. Cryst.* **25**, 495–503.
- Svergun, D. I. (1999). *Biophys. J.* **76**(6), 2879–2886.
- Svergun, D. I., Barberato, C. & Koch, M. H. J. (1995). *J. Appl. Cryst.* **28**, 768–773.
- Svergun, D. I. & Nierhaus, K. H. (2000). *J. Biol. Chem.* **275**(19), 14432–14439.
- Svergun, D. I., Petoukhov, M. V. & Koch, M. H. J. (2001). *Biophys. J.* **80**(6), 2946–2953.
- Svergun, D. I., Richard, S., Koch, M. H. J., Sayers, Z., Kuprin, S. & Zaccai, G. (1998). *Proc. Natl Acad. Sci. USA*, **95**(5), 2267–2272.
- Walther, D., Cohen, F. E. & Doniach, S. (2000). *J. Appl. Cryst.* **33**, 350–363.