

***PRIMUS*: a Windows PC-based system for small-angle scattering data analysis**

Petr V. Konarev, Vladimir V. Volkov, Anna V. Sokolova, Michel H. J. Koch and Dmitri I. Svergun

Copyright © International Union of Crystallography

Author(s) of this paper may load this reprint on their own web site provided that this cover page is retained. Republication of this article or its storage in electronic databases or the like is not permitted without prior permission in writing from the IUCr.

PRIMUS: a Windows PC-based system for small-angle scattering data analysis

Petr V. Konarev,^{a,b} Vladimir V. Volkov,^{a,b} Anna V. Sokolova,^{a,b} Michel H. J. Koch^b and Dmitri I. Svergun^{a,b*}

^aInstitute of Crystallography, Russian Academy of Sciences, Leninsky pr. 59, 117333, Moscow, Russia, and

^bEuropean Molecular Biology Laboratory, EMBL c/o DESY, Notkestrasse 85, D-22603, Hamburg, Germany.

Correspondence e-mail: svergun@embl-hamburg.de

A program suite for one-dimensional small-angle scattering data processing running on IBM-compatible PCs under Windows 9x/NT/2000/XP is presented. The main program, *PRIMUS*, has a menu-driven graphical user interface calling computational modules to perform data manipulation and analysis. Experimental data in binary OTOKO format can be reduced by calling the program *SAPOKO*, which includes statistical analysis of time frames, averaging and scaling. Tools to generate the angular axis and detector response files from diffraction patterns of calibration samples, as well as binary to ASCII transformation programs, are available. Several types of ASCII files can be directly imported into *PRIMUS*, in particular, sasCIF or ILL-type files are read without modification. *PRIMUS* provides basic data manipulation functions (averaging, background subtraction, merging of data measured in different angular ranges, extrapolation to zero sample concentration, etc.) and computes invariants from Guinier and Porod plots. Several external modules coupled with *PRIMUS* via pop-up menus enable the user to evaluate the characteristic functions by indirect Fourier transformation, to perform peak analysis for partially ordered systems and to find shape approximations in terms of three-parametric geometrical bodies. For the analysis of mixtures, *PRIMUS* enables model-independent singular value decomposition or linear fitting if the scattering from the components is known. An interface is also provided to the general non-linear fitting program *MIXTURE*, which is designed for quantitative analysis of multicomponent systems represented by simple geometrical bodies, taking shape and size polydispersity as well as interparticle interference effects into account.

© 2003 International Union of Crystallography
Printed in Great Britain – all rights reserved

1. Introduction

X-ray and neutron small-angle scattering (SAXS and SANS) are powerful techniques for the analysis of dispersed systems of different nature (inorganic, polymer, biological) (Feigin & Svergun, 1987). Recent decades have seen remarkable progress in X-ray and neutron scattering instrumentation (see e.g. Pedersen, 2002, for a review). High-flux dedicated X-ray synchrotron radiation and neutron beamlines can generate enormous amounts of data, e.g. in time-resolved experiments. Progress in the analysis of these data has been less dramatic, and the processing of small-angle scattering data remains to a large extent a non-trivial multistep task. Various data processing and manipulation packages exist at major large-scale facilities, but their specific input data formats often limit their portability between different laboratories. A non-exhaustive list of general data reduction systems includes e.g. the SANS package at HMI, Berlin (Keiderling, 1997), the *GRASP* package at the ILL, Grenoble (Ghosh, 1989; Dewhurst, 2002), and the data acquisition/reduction system at the DUBBLE beamline at the ESRF, Grenoble (Homan *et al.*, 2001). The analysis typically yields processed scattering patterns and a few overall parameters characterizing the system. Given the improved experimental capabilities and shorter

collection times on modern sources, there is a need for flexible general purpose data analysis tools. In the present paper, a menu-driven program package, *PRIMUS*, is presented, which is capable of reading different input file formats and allows the user to perform standard data reduction and manipulation tasks and computation of the invariants. Most notably, the package includes interfaces to several programs for the analysis of one-dimensional scattering data from monodisperse systems, polydisperse and interacting systems and mixtures, as well as partially ordered systems. This 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 graphical user interface (GUI). The program package can be implemented on IBM-compatible PCs under Windows 9x/NT/2000/XP.

2. Primary data processing in OTOKO format

The current version of *PRIMUS* deals with one-dimensional scattering data sets (e.g. those collected by linear detectors or radially averaged patterns). *PRIMUS* can either read in ASCII data files (see next section) or unprocessed binary data files in the OTOKO format

(Boulin *et al.*, 1986). The one-dimensional OTOKO data consist of an ASCII header and of two associated binary files, the first one containing the scattering patterns of successive time frames and the second one the corresponding ancillary data, such as ionization chamber readings, current in the storage ring, waiting and reading time, *etc.* OTOKO format is used at the X33 beamline of the EMBL on the storage ring DORIS of the Deutsches Elektronen Synchrotron (DESY) at Hamburg [where the data are generated by the acquisition program *SUPER_WIN* (Golding, 2002)], as well as at other facilities (HASYLAB, DESY, beamline A2; NSLS, Brookhaven, beamline X27C; LURE, Orsay, beamline D24; Daresbury Laboratory, several beamlines; SSRL, Stanford, beamline BL4-2). OTOKO files may be converted from Big Endian (UNIX) to Little Endian (PC) format and *vice versa* by the program *ENDCON*.

Data processing follows the standard equation (*e.g.* Strunz *et al.*, 2000) to yield the net scattering intensity $I(n)$ as a function of detector channel or averaging bin n :

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

where the subscripts s , m and e denote the scattering from sample, matrix (*e.g.* solvent) and empty cell (camera background), subscript 0 refers to the calibration channel reading corresponding to the intensity of the incident beam, T stands for transmission, c for the specimen concentration in the sample and $\text{Det}(n)$ is the detector response function. The intensities $I_s(n)$, $I_m(n)$, $I_e(n)$ are either numbers of events (photons/neutrons) registered by the detector or are averages obtained by radial binning and/or by summing up the results of repetitive experiments on the same sample. Depending on the type of system, some terms may be missing in equation (1): *e.g.* in

solution scattering T_s often equals T_m and the third term vanishes; for non-particulate systems, there is no matrix scattering ($T_m \rightarrow \infty$) and the second term is omitted.

Data reduction of OTOKO files is done by the program *SAPOKO* either interactively or through an ASCII batch file. The program performs several operations depending on the type of measurement. The data are first multiplied by a mask function to eliminate the unused channels. For time-resolved measurements the data from successive time frames are normalized against calibration channels (*e.g.* ionization chamber reading) and divided by the detector response. The ancillary data are stored in the second binary file and the experimental errors calculated from the numbers of counts in the unprocessed experimental data following standard statistical procedures (Bevington, 1969) in a third binary file. Static patterns are also recorded as a set of successive time frames to monitor possible radiation damage, and one of the frames, usually the first, is taken as a reference. After data masking and normalization, all frames in the file are compared with the reference one. They are accepted for averaging only if they are statistically compatible with the reference frame following χ^2 statistics (Bevington, 1969). The ancillary data are not kept for static patterns and the experimental errors are stored as the second record of the – in this case single – binary file.

PRIMUS provides a convenient menu interface to run *SAPOKO*, whereby the input parameters and the range of files to be processed are specified in a pop-up menu (displayed in Fig. 1 along with the overall GUI of *PRIMUS*). The detector response function for data calibration can be created in *PRIMUS* using a ‘Detector’ pop-up menu dialog. Experimental file(s) containing the patterns recorded for uniform detector illumination (*e.g.* radioactive source or fluorescence of a metal foil) are fitted by appropriate parametric functions

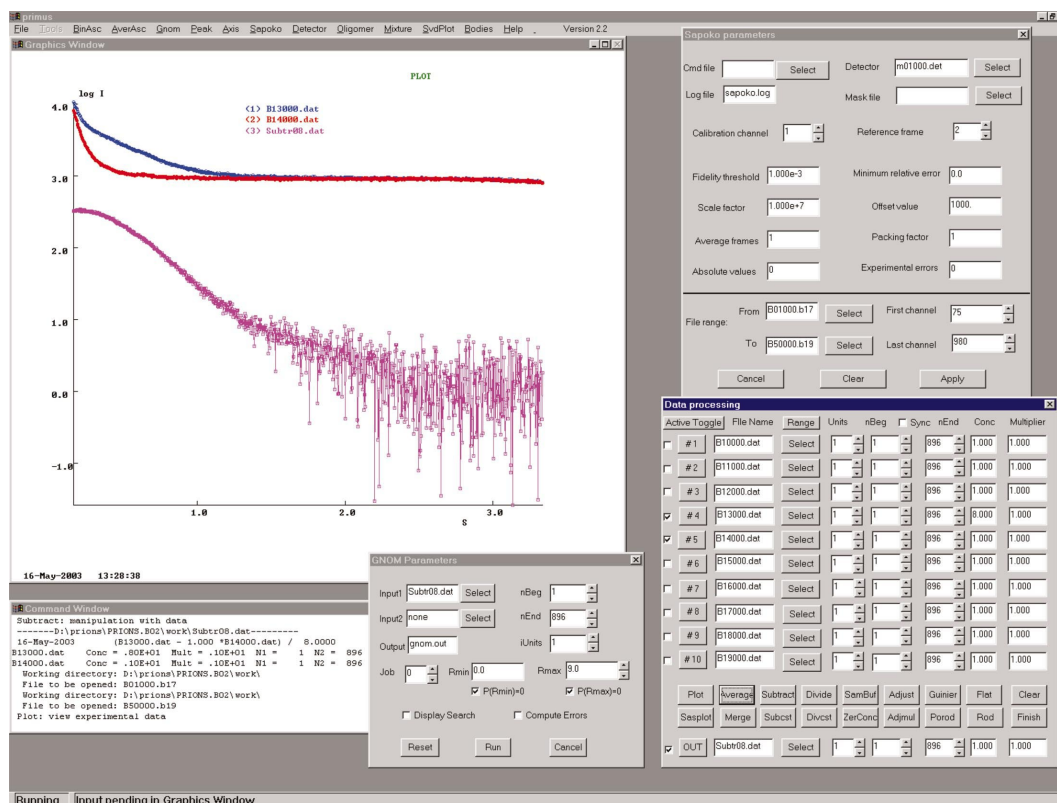


Figure 1 User interface of *PRIMUS* displaying the data processing toolbox, graphics and text windows. The blue curve is the scattering profile from a solution of bovine serum albumin, the red one that of the solvent and the magenta curve is the net scattering after background subtraction. Pop-up dialog boxes calling the primary data reduction program *SAPOKO* and the indirect transformation program *GNOM* are also displayed.

(Lorentz, exponential or polynomial). The ratio to the fit is used as the $\text{Det}(n)$ function in equation (1) to compensate for the non-uniformity of the detector response. The angular axis can be generated from the 'Makeaxis' dialog using patterns from standard samples with appropriate periodicities (e.g. dry collagen, tripalmitin or Ag behenate). The peaks selected interactively as described below for the program *PEAK* are used to establish the relation between detector channel number and momentum transfer, $s = 4\pi \sin \theta / \lambda$, where 2θ is the scattering angle and λ the wavelength. Using the axis information, the processed binary data are converted from OTOKO to ASCII format by using the options 'Binasc' (for multiple time-resolved frames) or 'Averasc' (for averaged static frames).

The program *SAPOKO* also includes options for background subtraction and scaling using equation (1), and given the scattering vector axis, can estimate overall particle parameters. There exists also a version of the indirect transformation program *GNOM* (Svergun, 1992) called *GNOMOKO*, which can directly use the output of *SAPOKO* to compute the characteristic functions. For further analysis of processed data it is convenient to use the extensive tools provided by *PRIMUS* for ASCII files, described in the following sections.

3. Data-manipulation toolbox

The previous section described an instrument-specific front end of *PRIMUS* where the raw data are reduced and converted into ASCII format. In the following, all manipulations are done with the ASCII data in tabular format containing the scattering vector or momentum transfer axis ' s ' in the first, the intensity ' $I(s)$ ' in the second and optionally the standard error $\sigma(s)$ in the third column. The program has a smart reader to recognize the beginning of the data stream in the input file, so that, e.g. sasCIF (Malfois & Svergun, 2000), ILL (Ghosh, 1989) or LOQ (King & Heenan, 1996) formatted files are read without modification. All valid data lines containing increasing values of momentum transfer and non-zero intensity are read and up to ten files can be loaded into the data-manipulation toolbox (filtering options are provided for multiple file selection). The user interface of *PRIMUS* displaying the toolbox is presented in Fig. 1. The angular axis in the input file can be given either in $s = 4\pi \sin \theta / \lambda$ or in $S = 2 \sin \theta / \lambda$ and the units can be transformed between nm^{-1} and \AA^{-1} as well as between s and S by a selection bar. The sample concentration and scale multiplier (default values are 1.0) as well as first and last data points can be adjusted for each data set individually. Each file can be made 'Active' (for plotting and other operations) by ticking the appropriate box (by default, newly loaded files are marked 'Active'). The standard data plotting window is displayed in Fig. 1 on the left; advanced plotting options are provided by calling the option 'Sasplot', which displays currently active files in a separate window using different scales and permits the user to select the plotting range and zoom factor with the mouse. The program *SASPLOTT* can also be run separately from outside *PRIMUS*.

The data-manipulation toolbox enables subtraction of the background, normalization of the data using equation (1), averaging, splicing, merging, etc. There are ten 'Input' file boxes and one 'Output' box, where the result of any operation is placed and can be further used in subsequent data manipulations. Linear operations with two files are 'Average', 'Subtract', 'Divide', as well as functions 'Adjmul' (least-squares scaling), 'Adjust' (scaling and constant adjustment). The last two functions also provide estimates of whether the two data sets are statistically different up to a scale factor (and, in the second case, also a constant term). Division by a factor and subtraction of a constant are performed with the 'Divcst' and 'Subcst' menu options. The option 'Merge' merges several files with auto-

mated least-squares scaling in the overlap(s) and possible joining of points in the output file. For all operations, information about the operation performed (type of operation, file names, weights, ranges of points used etc.) is written in the header of the output file and the headers of the operands are appended to the output data stream. The default resulting file name contains the operation name with incrementing number, and with these conventions, full data processing history can be tracked. The name in the 'Out' box file (or, if this box is empty, the name of the first active data set) is taken as the default input file name for all callable data analysis programs described below.

4. Data analysis with *PRIMUS*

4.1. Concentration series analysis

PRIMUS provides two options for data extrapolation to zero specimen (solute) concentration in the sample (e.g. protein concentration in solution). The first one, 'SamBuf', solves a system of linear equations for a series of measurements at different specimen concentrations assuming the experimental data to be a linear combination:

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

where v_k is the volume fraction of the specimen in the sample. Solution of the system separates two functions, $I_{\text{sample}}(s)$ and $I_{\text{buffer}}(s)$, which is useful for subtracting the background for measurements made at higher angles and high specimen concentrations, in which case the volume fraction of the specimen must be taken into account in equation (1). For proteins, the estimate of the volume fraction of the specimen is $v_k = 7.3 \times 10^{-4} c_k$, where c_k is the concentration in mg ml^{-1} , and $0.73 \text{ cm}^3 \text{ g}^{-1}$ is the average partial specific volume of a protein.

The second option, 'ZerConc', accounts for attractive or repulsive interactions, which mostly influence the scattering data at very small angles. For this, the scattering patterns after background subtraction are scaled such that they coincide at higher angles (where the structure factor may be neglected). The initial portion of the data is extrapolated to zero concentration assuming linear concentration dependence and the resulting intensities are merged with the scattering from the sample with the highest concentration at higher angles.

4.2. Computation of overall parameters

The radius of gyration R_g and forward scattering $I(0)$ are computed by selecting the 'Guinier' option using the Guinier approximation:

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

which is valid for $(sR_g) < 1.3$ (Guinier, 1939). The first and last points in the fit can be changed interactively by appropriate handlers and the Guinier plot is automatically redrawn. The 'Flat' and 'Rod' buttons can be used to compute the radius of gyration of the thickness of flat particles and the radius of gyration of the cross section of rod-like particles, respectively.

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

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

The angular range for computation can be adjusted interactively as described for the Guinier plot, whereby the constant K is auto-

matically calculated from the Porod asymptote [$I(s)s^4$ versus s^4] at higher angles.

Structural characteristics of partially ordered systems are computed from the maxima in the scattering profiles using the option 'Peak' from the main menu of *PRIMUS*. The default input file name for the *PEAK* module is transferred from the 'Tools' box as described above, and the fitting region(s) for a single or for multiple peaks are selected with the mouse. The marked peaks are fitted individually (either on linear or on logarithmic scale) by Gaussian functions, allowing also for two-parametric background subtraction, and the fits are displayed on the screen (see Fig. 2 for the *PEAK* interface). The structural parameters, which can be either displayed or stored, are defined by the following expressions, where λ represents the wavelength of the radiation, β_s is the full width at half-maximum of the peak (in radians) and θ_{\max} is the scattering angle corresponding to the value of the momentum transfer corresponding to the maximum (s_{\max}). The d spacing of the peak is $d = 2\pi/s_{\max}$; the long-range order dimension $L = \lambda/(\beta_s \cos \theta_{\max})$; the interaction radius $R_m = (\pi/2.5)^2(\lambda/\beta_s)$ and the degree of disorder $\Delta d = (1/\pi)(\beta_s d/\lambda)^{1/2}$ (Vainshtein, 1966). The capabilities of *PEAK* (which can also be run outside *PRIMUS*) are similar to the peak fitting options in the *FISH* system (Heenan, http://www.isis.rl.ac.uk/largescale/LOQ/FISH/FISH_intro.htm). The program *FISH* has wider fitting possibilities but does not provide an interactive peak-selection interface.

4.3. Calling external data analysis modules

For computation of characteristic functions of dilute monodisperse or polydisperse systems, *PRIMUS* provides an interface to the indirect transformation program *GNOM* (Semenyuk & Svergun, 1991; Svergun, 1992). This program solves an integral equation of the type

$$I_{\text{exp}}(s) = \int_{r_{\min}}^{r_{\max}} p(r) K(s, r) dr, \quad (5)$$

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 an integral kernel of the corresponding Fourier transformation. The user-defined parameters are specified in the dialog box which pops up when calling *GNOM* (Fig. 1). The default input file name is transferred from the

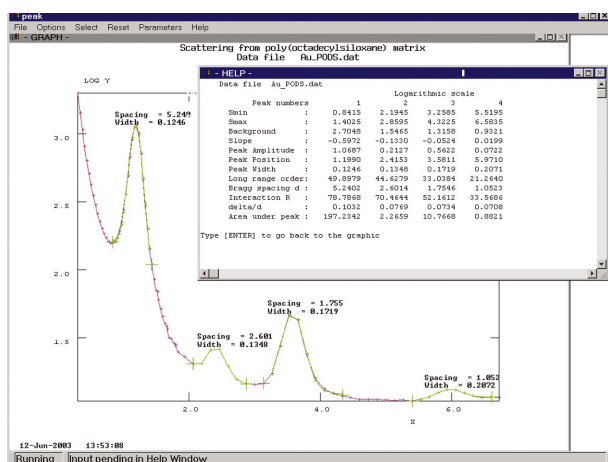


Figure 2 User interface of the *PEAK* module displaying the scattering pattern of a poly(octadecylsiloxane) polymer matrix with embedded metal nanoparticles (Svergun *et al.*, 2001). The experimental curve is drawn in magenta; the green curves are the fitted peak profiles. The corresponding structural parameters are displayed in the pop-up window.

'Tools' box as described above; if two input files are specified, they are automatically merged to correspond to the same characteristic function. The user may select the type of system (monodisperse or polydisperse), ranges in real and reciprocal space, etc. When started in the default mode, *GNOM* yields the distance distribution function of a particle given its maximum diameter r_{\max} without further user intervention (for more details about *GNOM* see Svergun, 1992).

Fast and simple particle shape analysis is performed by the program *BODIES*, which finds the best approximation of the initial portion of the scattering data in terms of three-parameter geometrical bodies. Approximate values of the particle volume, R_g and r_{\max} are estimated from the experimental data to select the initial approximation and fitting range. This range is automatically selected to contain about four Shannon channels (Shannon & Weaver, 1949), i.e. the data are fitted up to $s = 4\pi/r_{\max}$, accounting for the three parameters describing the model plus the scaling coefficient. The user may choose between six shapes (triaxial ellipsoid, ellipsoid of revolution, circular, elliptical or hollow cylinder, prism); by default, the program finds the best approximation in each class of bodies. The scattering intensities from the geometrical bodies $I(s)$ are evaluated using standard formulae (Feigin & Svergun, 1987). The non-linear least-squares package NL2SOL (Dennis *et al.*, 1981) is employed to find the geometrical parameters minimizing the discrepancy

$$\chi^2 = [1/(N - 1)] \sum_{j=1}^N \{[\mu I(s_j) - I_{\text{exp}}(s_j)]/\sigma(s_j)\}^2, \quad (6)$$

where N is the number of experimental points and μ is the scaling factor. The program displays and saves the appropriate fit(s) along with the parameters yielding the best approximation.

For model-independent analysis of multiple scattering data sets from polydisperse systems, singular value decomposition (SVD) (Golub & Reinsh, 1970) is implemented. The matrix $\mathbf{A} = \{A_{ik}\} = \{I^{(k)}(s_i)\}$, ($i = 1, \dots, N$, $k = 1, \dots, K$, where K is the number of data sets) is represented as $\mathbf{A} = \mathbf{U} \cdot \mathbf{S} \cdot \mathbf{V}^T$, where the matrix \mathbf{S} is diagonal, and the columns of the orthogonal matrices \mathbf{U} and \mathbf{V} are the eigenvectors of the matrices $\mathbf{A} \cdot \mathbf{A}^T$ and $\mathbf{A}^T \cdot \mathbf{A}$, respectively. The matrix \mathbf{U} yields a set of so-called left singular vectors, i.e. orthonormal basic curves $U^{(k)}(s)$, that spans the range of matrix \mathbf{A} , whereas the diagonal of \mathbf{S} contains their associated singular values in descending order (the larger the singular value, the more significant the vector). Physically, the number of significant singular vectors (non-random curves with significant singular values) yields the minimum number of independent curves required to represent the entire data set by their linear combinations (e.g. for mixtures, see below). The program *SVD PLOT* computes the SVD from the active data sets in the toolbox and displays the singular vectors and singular values (see example in Fig. 3). A non-parametric test of randomness due to Wald and Wolfowitz (Larson, 1975) is implemented to obtain the number of significant singular vectors, which provides an estimate of the minimum number of independent components in equilibrium or non-equilibrium mixtures [e.g. number of (un)folding or assembly intermediates].

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

For a system containing distinct components with known scattering intensities, the scattering pattern is a linear combination

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

where v_k and $I_k(s)$ are the volume fraction and the scattering intensity from the k th component, respectively. Given the experimental scattering from the mixture $I_{\text{exp}}(s)$ and scattering curves of the components, $I_k(s)$, the program *OLIGOMER* implements a non-negative linear least-squares algorithm (Lawson & Hanson, 1974) to find the volume fractions minimizing the discrepancy in equation (6). The program, which can be either launched from *PRIMUS* or run separately, is useful to characterize well defined systems like oligomeric equilibrium mixtures of proteins with known high-resolution structure.

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

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

where $\Delta\rho_k(R)$, $V_k(R)$ and $i_{0k}(s, R)$ denote the contrast, volume and normalized scattering intensity (form factor) 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 based on the scattering data. The algorithm allows one to model mixtures containing up to ten different components (*i.e.* different types of particles), each described by its volume fraction, form factor, contrast, polydispersity and, for spherical particles, potential for interparticle interactions. 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 $R_k^{(s)}$]. Each component can be monodisperse or polydisperse, and in the second case the size polydispersity $D_k(R)$ is described by a two-parametric monomodal Gaussian or Schultz distribution characterized by the average dimension R_{0k} and dispersion ΔR_k . Interparticle interactions can be taken into account within each component consisting of spherically symmetric particles. The structure factor $S_k(s)$ is represented in the Percus–Yevick approximation using the sticky hard-sphere potential

(Baxter, 1968) described by two parameters: the hard-sphere interaction radius R_k^{hs} and the ‘stickiness’ τ_k .

Non-linear fitting in *MIXTURE* is based on the program suite *OPTIS* that encompasses more than a dozen powerful optimization methods (Volkov, 1995). By default, the program uses a quasi-Newton minimization procedure due to Broyden–Fletcher–Goldfarb–Shanno with simple bounds on the variables (Gill *et al.*, 1981), which allows physically justified limitations to be placed on the parameters. The use of *MIXTURE* requires a command file 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, type of distribution function (Gauss or Schultz), *etc.* The upper and lower limits can be specified for all fitting parameters. An example of the format of the command file for the system of polydisperse solid spheres and cylinders is presented in Table 1. Each adjustable parameter has the following format: initial value, lower boundary limit, upper boundary limit. If the three values are equal, the value of the parameter will be fixed.

Instead of *EXPERIMENT*, one can use the *TEST* mode to work with simulated data, which is useful for model calculations (*e.g.* for checking the stability of solution provided by *MIXTURE*). One can also design a model containing several components of the same type (for example large and small spheres, long and short cylinders, *etc.*), in which case the components must be separated in size by initial approximation and boundary limits.

Currently, *MIXTURE* is run from *PRIMUS* by a pop-up menu prompting for the command file and experimental data file names (the interface is being further developed to allow the user to enter and/or modify the information about the components interactively). After non-linear minimization, the 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 record the history of running *MIXTURE* under different conditions.

A prototype of *MIXTURE* was developed in the study of AOT water-in-oil microemulsions and applied for quantitative description

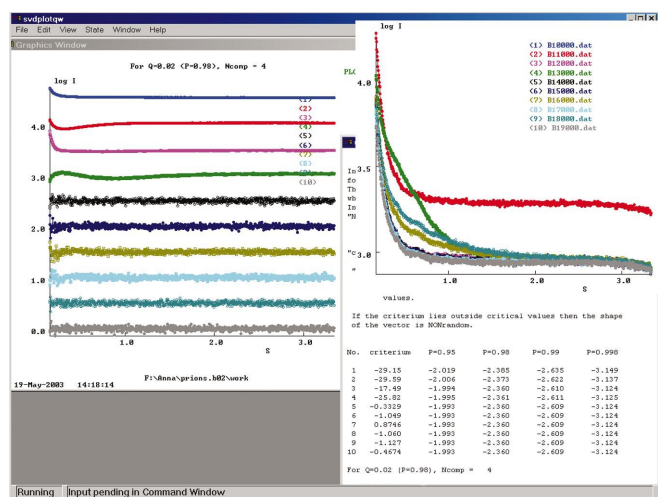


Figure 3 Example of SVD of ten data sets containing three different protein solutions, six buffers and the scattering from an empty cell. The initial data are displayed in the inset (top right corner); the singular vectors are shown in the *SVD PLOT* graphics window on the left. The statistical test yields four independent singular vectors.

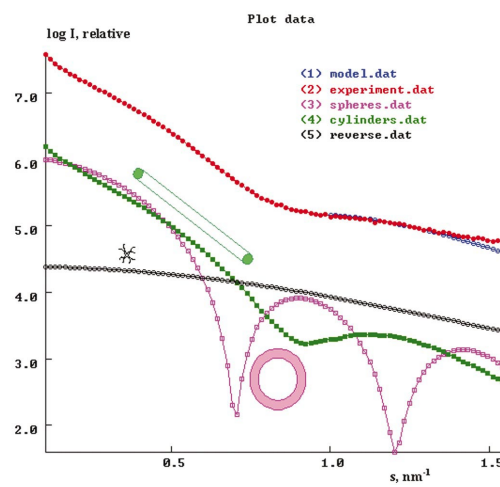


Figure 4 Example of data fitting by the scattering from a three-component system for AOT/water/oil microemulsions. The experimental curve (2) is fitted by the total model intensity (1) represented by a linear combination of contributions from spherical droplets (3, magenta), long cylinders (4, green) and reversed micelles (5, black). The shapes of the three components are schematically displayed on the plot made by *PRIMUS*.

Table 1

Example of a *MIXTURE* command file for a two-phase system.

Test command file	Comment line 1
Spheres + Cylinders	Comment line 2
EXPERIMENT	Data fitting mode
2	Number of components
0.20	Total volume fraction
SPHERE	Type of the first component
0.25, 0.0, 1.0	Sphere volume fraction†
0.0, 0.0, 0.0	Inner shell radius†
0.0, 0.0, 0.0	Inner shell contrast†
48.0, 40.0, 60.0	Outer shell radius†
1.0, 1.0, 1.0	Outer shell contrast†
5.0, 0.1, 25.0	Sphere polydispersity†
75.0, 55.0, 95.0	Hard-sphere radius†
2	Polydispersity (1 = Gauss, 2 = Schultz)
0.0, 0.0, 0.0	Sticky parameter (0 = hard sphere)†
CYLINDER	Type of the second component
0.25, 0.0, 1.0	Cylinder volume fraction†
0.0, 0.0, 0.0	Inner shell radius†
0.0, 0.0, 0.0	Inner shell contrast†
40.0, 30.0, 60.0	Outer shell radius†
1.0, 1.0, 1.0	Outer shell contrast†
3.0, 2.1, 10.1	Radius polydispersity†
2	Polydispersity (1 = Gauss, 2 = Schultz)
300.0	Cylinder length
2	1 = OTOKO data, 2 = ASCII data
test1.dat	Experimental data file
1.0	Angular scale multiplier

† Adjustable parameter (see §5).

of the droplet to cylinder transition in these classical microemulsion systems (Svergun *et al.*, 2000). The system was found to be a mixture of three components: (i) spherical water droplets, moderately polydisperse, average radius depending on the water/AOT ratio (w_0), (ii) long cylindrical aggregates, (iii) reverse AOT micelles containing bound water (an example of scattering from individual components and from the entire mixture is shown in Fig. 4). The present version of *MIXTURE* has been rewritten, further generalized and coupled with *PRIMUS* (although it is also possible to run the program from a console prompt). The features offered by *MIXTURE* are complementary to those provided by the *FISH* package: while the latter offers a wider selection of form and structure factors, *MIXTURE* is able to handle systems consisting of several distinct components.

6. Conclusions

The software packages described here are intended for use with one-dimensional patterns. Two-dimensional data processing programs are available from several sources (*e.g.* Hammersley, 1995; Stribeck, 2001). Our aim is to develop a software system allowing the comprehensive and convenient processing of one-dimensional X-ray and neutron patterns for different types of isotropic and partially ordered systems, like solutions of biopolymers and colloidal particles, microemulsions, polymers in solution and in bulk. The current functionality of *PRIMUS* is complementary to that of the packages described by Dewhurst (2002), Homan *et al.* (2001) and Hiragi *et al.* (2003). The executable code of *PRIMUS* runs on any Windows PC and does not require any additional software packages or languages like Matlab, IgorPro or IDL. *PRIMUS* and all the modules called by it have extensive on-line help invoked for each program by using the 'Help' menu option. The program is readily extendable to call other data processing and analysis programs and it is planned to do so in the future, based on feedback and requests from users. In particular, it is planned to extend the range of input formats and to implement advanced modelling of the scattering data where the sample and background data are separately submitted to the fitting programs to omit the background subtraction step.

The programs *PRIMUS*, *SASPLOTT*, *PEAK*, *SAPOKO*, *OLIGOMER*, *BODIES*, *SVDPLOTT* and *MIXTURE* have been installed on the Windows PC used for data collection at the X33 beamline of the EMBL and extensively tested on a number of ongoing projects at the EMBL, Hamburg Outstation, and in the Institute of Crystallography, Moscow. All these programs as well as other packages for advanced analysis of small-angle scattering data (executable codes and user instructions) are available from the Web site of the EMBL outstation at DESY, Hamburg (<http://www.embl-hamburg.de/ExternalInfo/Research/Sax/index.html>).

The work was supported by INTAS grant 00-243, DAAD grant (2002) to PVK and INTAS grant YSF 2001/2-0133 to AVS.

References

- Baxter, R. J. (1968). *J. Chem. Phys.* **49**, 2270–2273.
- Bevington, P. B. (1969). *Data Reduction and Error Analysis for the Physical Sciences*. New York: McGraw-Hill.
- Boulin, C., Kempf, R., Koch, M. H. J. & McLaughlin, S. M. (1986). *Nucl. Instrum. Methods A*, **249**, 399–407.
- Dennis, J., Gay, D. & Welsch, R. (1981). *ACM Trans. Math. Soft.* **7**, 348–383.
- Dewhurst, C. (2002). *Graphical Reduction and Analysis SANS Program for Matlab*, http://www.ill.fr/lss/grasp/grasp_manual.pdf.
- Feigin, L. A. & Svergun, D. I. (1987). *Structure Analysis by Small-Angle X-ray and Neutron Scattering*. New York: Plenum Press.
- Gill, P. E., Murray, W. & Wright, M. H. (1981). *Practical Optimization*. London: Academic Press.
- Ghosh, R. E. (1989). *A Computing Guide for Small-Angle Scattering Experiments*. Institut Laue-Langevin Internal Publication 89GH02T, Grenoble.
- Golding, F. (2002). Personal communication.
- Golub, G. H. & Reinsh, C. (1970). *Numer. Math.* **14**, 403–420.
- Guinier, A. (1939). *Ann. Phys. (Paris)*, **12**, 161–237.
- Hammersley, A. P. (1995). *FIT2D: V5.18 Reference Manual V1.6*, ESRF Internal Report, Exp/AH/95-01, Grenoble.
- 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.
- King, S. M. & Heenan, R. K. (1996). *LOQ Instrument Handbook*, Rutherford Appleton Laboratory Internal Publication RAL-TR-96-036.
- Konarev, P. V., Petoukhov, M. V. & Svergun, D. I. (2001). *J. Appl. Cryst.* **34**, 527–532.
- Larson, H. J. (1975). *Statistics: an Introduction*, pp. 328–329. New York: John Wiley.
- Lawson, Ch. 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.
- Pedersen, J. S. (2002). *Neutrons, X-rays and Light: Scattering Methods Applied to Soft Condensed Matter*, edited by P. Lindner & Th. Zemb, pp. 127–144. Amsterdam: North Holland Delta Series.
- Porod, G. (1982). *General Theory*, in *Small-Angle X-ray Scattering*, edited by O. Glatter & O. Kratky. London: Academic Press.
- 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.
- Stribeck, N. (2001). *J. Appl. Cryst.* **34**, 496–503.
- Strunz, P., Saroun, J., Keiderling, U., Wiedenmann, A. & Przenioslo, R. (2000). *J. Appl. Cryst.* **33**, 829–833.
- Svergun, D. I. (1992). *J. Appl. Cryst.* **25**, 495–503.
- 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**, 651–1665.
- Svergun, D. I., Shtykova, E. V., Kozin, M. B., Volkov, V. V., Konarev, P. V., Dembo, A. T., Shtykova, E. V. Jr, Bronshtein, L. M., Chernyshov, D. M., Platonova, O. A., Yakunin, A. N., Valetsky, P. M. & Hohlov, A. R. (2001). *Crystallogr. Rep.* **46**, 586–595.
- Vainshtein, B. K. (1966). *Diffraction of X-rays by Chain Molecules*. Amsterdam: Elsevier.
- Volkov, V. V. (1995). Unpublished results.