Modelling of multicomponent polydisperse systems using small-angle scattering data

P V Konarev1,2, A E Kryukova1,2, V V Volkov1 and D I Svergun3
1Laboratory of Reflectometry and Small-angle Scattering, A.V. Shubnikov Institute of Crystallography of FRC "Crystallography and photonics", Russian Academy of Sciences, Leninsky pr. 59, 119333, Moscow, Russia
2NRC "Kurchatov Institute", pl. Kurchatova 1, 123098 Moscow, Russia
3European Molecular Biology Laboratory (EMBL), Hamburg Outstation, Notkestrasse 85, 22607, Hamburg, Germany

E-mail: konarev@ns.crys.ras.ru

Abstract. Algorithms for modelling of multicomponent polydisperse systems using small-angle scattering data have been developed. They cover a wide range of particle form-factors and take into account interparticle interactions. The algorithms are implemented in computer programs MIXTURE and POLYSAS, and their performance is tested on a number of simulated and experimental examples. The stability of the solution search is explored for the case of a two-component polydisperse system of spherical particles.

1. Introduction

Structural characterization of nanoscale objects is important for a better understanding of various physical, chemical and biological systems. The methods like X-ray crystallography, nuclear magnetic resonance and electron microscopy provide valuable information about particle shape with high resolution up to a few Å. However, these methods need special sample preparations, which are not always easy, and upon crystallization, freezing or isotope labelling artefacts or differences compared to the native state of the object can be observed. In contrast, small-angle scattering (SAS) does not require special treatment of the sample and permits to study materials in their native conditions yielding particle envelopes at low resolution of 1-2 nm [1]. The use of SAS in combination with other structural methods yields complementary information and improves the reliability of the structural results.

Depending on the level of structural organization one can divide the nanostructured systems into the following categories: monodisperse systems consisting of identical non-interacting particles (e.g. diluted solutions of macromolecules, metal nanoparticles, self-assembled polymer clusters etc.), polydisperse multicomponent systems containing interacting particles of different sizes and shapes (mixtures of proteins, nanoparticles and microemulsions) and partially oriented systems (liquid crystals, crystallized polymers etc.). For monodisperse objects, SAS allows one to determine the geometrical characteristics (radius of gyration, maximum size, excluded volume of the particle) and particle shape. For multicomponent systems the main structural task of SAS analysis is to determine the number and the type of components, and estimate their volume fractions, average sizes and polydispersity. For the partially oriented systems, the scattering patterns contain diffraction peaks, and
their positions and widths determine the lattice plain distances, the sizes of crystallites and the degree of disorder.

During last decades effective algorithms on low resolution ab initio shape determination for monodisperse solutions have been developed and are widely used [2]. Unfortunately, they cannot be directly applied for polydisperse multicomponent systems since the simultaneous restoration of the particle shape and the size distributions is a highly ambiguous task. To overcome this complexity the shape of the particle is often approximated by simple geometrical body (sphere, cylinder, ellipsoid etc.) and the size distribution functions of the particles can be restored in this case.

In the present paper we describe the functionality and potential applications of the program MIXTURE [3] and its interactive graphical version implemented in POLYSAS [4] designed to analyze multicomponent systems containing various types of polydisperse particles and take into account the interparticle interactions. These programs belongs to the package ATSAS [2] freely available to the academic users. The performance of the programs will be illustrated by several practical examples. The tests of the stability of the restored solutions for two-component systems will be presented and the improvements in the search algorithms will be discussed.

2. Modelling of polydisperse mixtures using small-angle scattering data

The scattering intensity from a system containing distinct components is a sum of the corresponding intensities which are weighted by their volume fractions:

\[ I(s) = \sum_{k=1}^{N} v_k I_k(s), \]

where the sum goes over different N components, \( v_k \) and \( I_k(s) \) are the volume fraction and the scattering intensity from the \( k \)-th component, respectively, \( s \) is the momentum transfer \( s = (4\pi/\lambda)\sin(\theta) \), where \( 2\theta \) is the scattering angle, \( \lambda \) is the X-ray or neutron wavelength.

Each component (e.g. particles of certain shape) in a dilute mixture can have its own polydispersity in particle size and certain type of interparticle interactions, and its intensity in the general case can be calculated according to the following expression:

\[ I_k(s) = S_k(s) \int D_k(R)V_k(R)[\Delta \rho_k(R)]^2 i_{sk}(s,R)dR, \]

where \( D_k(R) \), \( V_k(R) \) and \( i_{sk}(s,R) \) denote the volume size distribution of the particles, the effective volume of the particle with size \( R \) and the normalized form-factor for the \( k \)-th component, respectively. Here, \( \Delta \rho_k(R) \) is the scattering contrast of \( k \)-th component and \( S_k(s) \) its structural factor, the latter responsible for the interparticle interactions (for dilute systems \( S_k(s) \) is simply equal to 1).

The use of equations (1) and (2) permits one to analyze complicated cases of polydisperse systems using SAS data. A program MIXTURE [3] was developed to characterize mixtures of particles with different shapes (using geometrical bodies or high resolution models of macromolecules) containing up to ten components (types of particles).

Presently, MIXTURE contains a wide range of form-factors for the components, different versions of the program including the following types of particles: solid, hollow and core-shell spheres, solid, hollow and core-shell cylinders, solid, hollow and core-shell dumb-bells and dimers of dumb-bells, two-axes ellipsoids, high resolution models of macromolecules supplied by the user, bilayer lipid shells, mixtures of multilayer spherical lipid vesicles. Several components of the same type are allowed, e.g. one can design a system containing small and large spherical particles, long and short cylinders etc. Each component is characterized by its relevant structural parameters (e.g. the average size \( R \), the polydispersity degree \( dR \) from Gauss or Schulz type of distributions, the volume fraction etc.) and the non-linear minimization procedures (e.g. quasi Newton method due to Broyden-Fletcher-Goldfarb-Shanno [5]) with simple bounds on the variables are employed to find the mixture with the best overall fit to the experimental data.

MIXTURE can be run automatically using a command file with model specifications and initial values of the parameters as well as their upper and lower limits. Alternatively, POLYSAS provides a
convenient interface where MIXTURE can be launched from the pop-up menus and each parameter of the model can be changed interactively using sliders or spin buttons. After the minimization, the best fit to the experimental data and the partial intensities from each components are stored as separate files, and the optimized parameter values are written in a log file after each run to build history of running MIXTURE under different conditions.

3. Practical applications to proteins and nanoparticles

After validation using simulated data, MIXTURE was applied to a number of experimental X-ray data sets. Some examples are presented below to illustrate the capacity of the method to quantitatively describe the structural composition of multicomponent systems. The X-ray synchrotron scattering data were recorded in collaborative user projects on the X33 beamline of the EMBL [6] on the storage ring DORIS-III (DESY, Hamburg). Figure 1 presents the X-ray scattering data from native and mutant forms of lumazine synthase from *Aquifex aeolicus* (LASQ) and *Bacillus subtilis* (LSBS) [7]. The latter enzyme catalyzes the reaction of riboflavin biosynthesis (vitamin B$_2$), an essential element for living organisms. In solution it forms small and large icosahedral capsids with a diameter of 16 and 30-32 nm, respectively. The relative abundance of small and large capsids strongly depends on buffer composition and pH. Besides, incomplete and deformed capsids can also be present in solution. MIXTURE permitted to estimate volume fractions and polydispersity degree of above components at different conditions, the obtained results being in agreement with cryo-EM measurements [7] and indicate that the ability of lumazine synthase to form multiple assembly states has an important biological and physiological meaning in catalytic processes of riboflavin biosynthesis.

**Figure 1** Experimental SAXS data (dots) and MIXTURE fits (solid lines) of lumazine synthase. (1) LSBS in borate buffer (pH 7); (2-4) LASQ in phosphate buffer (pH 6,7,8); (5-7) LASQ in Tris buffer (pH 7,8,9); (8) LASQ mutant in Tris buffer (pH 8).

**Figure 2** SAXS scattering patterns (dots with experimental errors bars) and fits from MIXTURE (solid lines) for iron oxide nanoparticles (NP) (curve 1) and NPs embedded into viral capsid of brome mosaic virus (BMV) (curve 2).

Figure 2 displays the X-ray scattering data from iron oxide nanoparticles (NP) embedded into the viral capsid of brome mosaic virus (BMV) [8]. The FeO/Fe$_2$O$_3$ nanoparticles prepared of iron oleate
by thermal decomposition were stabilized by the coating with poly(maleic anhydride-alt-1-octadecene) (PMAOD) modified with poly (ethylene glycol) (PEG) tails. The encapsulation of NPs into the virus shells BMV was reached by mixing and subsequent dialysis of the solution. The experimental data were fitted by MIXTURE with a model containing mixtures monomeric virus-like nanoparticles (VNP), dimeric and tetrahedron-like “associates” of VNPs that can be seen from EM experiments [8]. It was found that the functionalized NPs tend to form significant fraction (around 50%) of these associates, while the BMV VNPs stay largely (80% volume fraction) monomeric. This result can be explained either by the screening of attractive magnetic NP interactions with the BMV shells or by the limited PEG tails penetration into the shells. The average distance between the individual NPs inside the clusters (around 28 nm) exceeds the NP sizes (25 nm) due to the PEG tails. The formation of mainly single particles for BMV VNPs is important for development of bioprobes with predicted and well-defined properties.

Both these examples demonstrate the capability of the program MIXTURE in elucidating the structural organization of multicomponent polydisperse systems.

4. Stability boundaries for the parameters of the restored solutions for two component systems of spherical particles

For complicated multi-component systems, ambiguous results may be obtained largely depending on the initial values of the parameters for the minimization. The stability of restored solutions using MIXTURE was assessed for a two-component systems of spherical particles [9, 10] with two well-separated size distributions at different noise levels. It was shown that the presence of moderate noise of Poisson type (up to 25%) does not significantly worsen the ranges of solution stability for the model parameters. At the same time the method may be highly sensitive to each individual pair of model parameters, and the contour plots showing the areas of finding the correct solution have complicated fractal-like boundaries with separated “islands” of instabilities. The latter ones can be explained both by the multimodal behaviour of the minimized functional and by the internal ambiguity of the given structural model. Similar trends were observed for the cases with partially overlapped size distributions. Three model systems were selected, which contain two types of spherical nanoparticles characterized by their radii $R_1$ and $R_2$, polydispersities $dR_1$ and $dR_2$ and volume fractions $v_1$ and $v_2$ (see Table 1).

| №, model number | $v_1$, relative | $v_2$, relative | $R_1$, nm | $R_2$, nm | $dR_1$, nm | $dR_2$, nm |
|-----------------|----------------|----------------|-----------|-----------|-----------|-----------|
| 1               | 0.75           | 0.25           | 6.0       | 8.0       | 0.1       | 1.0       |
| 2               | 0.75           | 0.25           | 7.0       | 8.0       | 0.3       | 0.3       |
| 3               | 0.50           | 0.50           | 6.0       | 8.0       | 0.7       | 0.7       |

The SAS curves for these systems without noise and with added Gaussian and Poisson noise (at a relative level of 25%) are displayed in Figure 3 (left panel). The corresponding volume distributions ($D_v(R)$ functions) are shown in Figure 3 (right panel). The contour plots of solution stability ranges versus the pair of parameters ($R_1, R_2$) presented in Figure 4 reveal that a wider range for the initial guess of the parameters can be chosen in the case of partially overlapped size distributions compared to the systems with two well-separated size distributions.

An improved search strategy may thus be suggested based on the above results. In future, such analysis will be expanded to real experimental data to help defining distinct model cases with higher probabilities for finding the ‘true’ solution for the system.
Figure 3. The left panel contains the simulated theoretical SAS curves with the parameters from Table 1 (curve 1 - with relative noise level 0%, curve 2 - Gaussian noise 25%, curve 3 - Poisson noise 25%). The curves are vertically displaced for better visualization. The right panel displays the volume size distribution functions $D_V(R)$.

Figure 4. The contour plots of solution stability ranges for three model systems depending on the pair of parameters $(R_1, R_2)$. The left column corresponds to the model 1, the central column – the model 2, the right one – the model 3. The plots for the curves without noise (relative noise level 0%) are located in the first row, with Gaussian noise (relative noise level 25%) - in the second row, with Poisson noise (relative noise level 25%) - in the third row. The interception of the dotted lines corresponds to the theoretical (‘true’) solution. If the restored parameters are located within 5% in the vicinity of ‘true’ solution, then the solution is treated as successful (red areas), otherwise the solution is treated as false (blue areas).
5. Conclusions
A program MIXTURE designed for the structural characterization of multicomponent polydisperse systems using SAS data was developed and successfully applied to a wide range of systems, including microemulsions [11], biological macromolecules [7] and nanoparticles [8]. The analysis of solution stability using the program on two-component systems [9, 10] showed a strong dependence of finding the ‘true’ solution on the type of the model parameters and on the experimental noise level and provided hints for the improved search strategy.

Acknowledgements
The authors acknowledge the support from the Federal Agency of Scientific Organizations (Agreement No 007-ГЗ/3363/26).

References
[1] Feigin L A and Svergun D I Structure Analysis by Small-Angle X-Ray and Neutron Scattering 1987 (New York: Plenum Press)
[2] Franke D, Petoukhov M V, Konarev P V, Pankovich A, Tuukkanen A, Mertens H D T, Kikhney A G, Hajizadeh N R, Franklin J M, Jeffries C M and Svergun D I 2017 J. Appl. Cryst. 50 1212-25
[3] Konarev P V, Volkov V V, Sokolova A V, Koch M H J and Svergun D I 2003 J. Appl. Cryst. 36 1277-82
[4] Konarev P V, Volkov V V and Svergun D I 2016 J. Phys. Conf. Ser. 747 012036
[5] Gill P.E., Murray W., Wright M.H. Practical Optimization 1981 (London: Academic Press)
[6] Blanchet C E, Zozulya A V, Kikhney A G, Franke D, Konarev P V, Shang W, Klaering R, Robrahm B, Hermes C, Ciprani F, Svergun D I and Roessle M 2012 J. Appl. Cryst. 45 489-95
[7] Zhang X, Konarev P V, Petoukhov M V, Svergun D I, Xing L, Cheng R H, Haase I, Fischer M, Bacher A, Ladenstein R and Meining W 2006 J. Mol. Biol. 362 753-70
[8] Malyutin A G, Cheng H, Sanchez-Felix O R, Carlson K R, Stein B D, Konarev P V, Svergun D I, Dragnea B and Bronstein L M 2015 ACS Appl. Mater. Interfaces 7 12089-98
[9] Kryukova A E, Konarev P V and Volkov V V 2017 J. Phys. Conf. Ser. 941 012069
[10] Kryukova A E, Konarev P V and Volkov V V 2018 Cryst. Reports 63 26-31
[11] Svergun D I, Konarev P V, Volkov V V, Koch M H J, Sager W F C, Smeets J and Blokhuis E M 2000 J. Chem. Phys. 113 1651-65