Form and structure factor, interactions, modelling

Petr V. Konarev

1 A.V. Shubnikov Institute of Crystallography, Federal Scientific Research Center “Crystallography and photonics” Russian Academy of Sciences, Moscow, Russia

2 National Research Center “Kurchatov Institute”, Moscow, Russia
SAXS experimental facilities in Russia

Laboratory setup

Synchrotron beamline “BioMUR” (Kurchatov Institute, Moscow) (built from former X33) in operation from Dec 2017
Outlines

Parametric modelling using least-squares methods and information content of SAS data

Form factors of simple geometrical bodies (spheres, cylinders, spherical core-shells, ellipsoids etc.)

Concentration effects, interactions and structure factors

Polydisperse & interactive systems in ATSAS
   Equilibrium oligomeric mixtures (OLIGOMER)
   Assembly/disassembly processes (SVDPLOT, MIXTURE)
   Restoring intermediates in evolving systems (DAMMIX)
   Graphical package for interactive processing (POLYSAS)
   Dissociation processes (GASBORMX, SASREFMX)
   Ensemble characterization of flexible systems (EOM)
Least-squares methods and parametric modelling

Experimental data:

Azimuthally averaged intensities

\[(s_i, I_{exp}(s_i), \sigma(s_i)), i=1, N\]

Data calculated from the model described by parameters \(\{a_j\}, j=1, M\)

\[(s_i, I_{mod}(s_i)), i=1, N\]

\[s_i = \frac{4\pi \sin(\theta_i)}{\lambda}\]

Fit quality

\[\chi^2 = \frac{1}{N-1} \sum_{i=1}^{N} \left( \frac{I_{exp}(s_i) - I_{mod}(s_i)}{\sigma(s_i)} \right)^2\]

\(\chi^2 = 1\) for \(N>>M\) corresponds to \(|I_{exp}(s_i) - I_{mod}(s_i)| = \sigma(s_i)\)

means statistical agreement between model and data

Information content of SAS data
Shannon sampling theorem: the scattering intensity from a particle with the maximum size \( D_{\text{max}} \) is defined by its values on a grid \( s_k = k\pi/D_{\text{max}} \) (Shannon channels):

\[
sI(s) = \sum_{n=1}^{\infty} s_n a_n \left[ \frac{\sin D_{\text{max}}(s - s_n)}{D_{\text{max}}(s - s_n)} - \frac{\sin D_{\text{max}}(s + s_n)}{D_{\text{max}}(s + s_n)} \right]
\]

Shannon sampling was utilized by many authors (e.g. Moore, 1980). An estimate of the number of channels in the experimental data range \( (N_s = s_{\text{max}} D_{\text{max}}/\pi) \) is often used to assess the information content in the measured data.
Given a (noisy, especially at high angles) experimental data set, which part of this set provides useful information for the data interpretation?

A usual practice is to cut the data beyond a certain signal-to-noise ratio but

• there is no objective estimation of the threshold
• this cut-off does not take into account the degree of oversampling
Application of sampling theorem to small-angle scattering data from monodisperse systems

Due to a finite experimental angular range, the data can be approximated by a truncated Shannon expression

\[ sI(s) \approx U_M(s) = \sum_{n=1}^{M} s_n a_n \left[ \frac{\sin D_{\text{max}}(s-s_n)}{D_{\text{max}}(s-s_n)} - \frac{\sin D_{\text{max}}(s+s_n)}{D_{\text{max}}(s+s_n)} \right] \]

\[ p(r) \approx p_M(r) = 8\pi r \sum_{n=1}^{M} s_n a_n \sin(s_n r) \]

The best approximation should minimize the discrepancy

\[ \chi^2(M) = \sum_{i=1}^{N} \frac{1}{2s_i^2 \sigma_i^2} [s_i I(s_i) - U_M(s_i)]^2 \]
Interpolation with different number of Shannon channels

Ellipsoid with half-axes 1, 15, 15 nm contains $M=38$ channels within angular range up to $s=4$ nm$^{-1}$
Interpolation with different number of Shannon channels

Ellipsoid with half-axes 1, 15, 15 nm contains M=38 channels within angular range up to s=4 nm$^{-1}$
Interpolation with different number of Shannon channels

$$\Omega(p) = \int_0^{D_{\text{max}}} \left[ \frac{dp_M(r)}{dr} \right]^2 dr$$

$$f(M) = \chi^2(M) + \alpha \Omega(p_M)$$

$$\alpha = \chi^2(M_{\text{max}}) / \Omega(p(M_{\text{min}}))$$
1. Automatically estimate $D_{\text{max}}$ (using AutoRg and AutoGnom (Petoukhov et.al., 2007))

2. Calculation of the nominal number of Shannon channels $N_S = s_{\text{max}} \pi/D_{\text{max}}$ and set up the search range $[M_{\text{min}};M_{\text{max}}]$, where $M_{\text{min}} = \max(3, 0.2*N_S)$, $M_{\text{max}} = 1.25*N_S$

3. For $M_{\text{min}} < M < M_{\text{max}}$, calculate the coefficients of Shannon approximation $a_n$ ($n=1,\ldots,M$) by solving system of equations using a non-negative linear least-squares procedure (Lawson & Hanson, 1974).

4. For each Shannon fit, calculate the discrepancy $\chi^2(M)$ and the integral derivative $\Omega(p_M)$.

5. Evaluate the scaling coefficient $\alpha$ as the ratio between $\chi^2(M_{\text{max}})$ and $\Omega(p(M_{\text{min}}))$

$$\Omega(p) = \int_{0}^{D_{\text{max}}} \left[ \frac{dp_M(r)}{dr} \right]^2 dr$$

$$f(M) = \chi^2(M) + \alpha \Omega(p_M) \quad \alpha = \frac{\chi^2(M_{\text{max}})}{\Omega(p(M_{\text{min}}))}$$

6. Determine the optimum value $M_S$ corresponding to the minimum of the target function $f(M)$

Examples of practical applications: **SAXS data** (Importin $\alpha/\beta$)

Importins $\alpha$ и $\beta$ mediate the import of nucleoplasmins through the nuclear pore, the latter ones interact with histones regulating the formation and shape of nucleosome.

*Shanum* estimates the effective number of Shannon channels $M=8$ and thus determines the useful angular range up to $s=1.3$ nm$^{-1}$.

Intensity from a system of monodisperse particles

\[ \frac{d\sigma(s)}{d\Omega} = I(s) = n\Delta\rho^2V^2P(s)S(s) = cM\Delta\rho_m^2P(s)S(s) \]

Number of scattered neutrons or photons per unit time, relative to the incident flux of neutron or photons per unit solid angle at s per unit volume of the sample

where

- \( n \) - the number density of particles
- \( \Delta\rho \) - the excess scattering length density given by electron density differences
- \( V \) - volume of the particles
- \( P(s) \) - the particle form factor, \( P(s=0)=1 \)
- \( S(s) \) - the particle structure factor, \( S(s=\infty)=1 \)

\( V \propto M \)
\( n = c/M \)
\( \Delta\rho \) can be calculated from partial specific density, composition
Form factor of a solid sphere

\[ I_{\text{sphere}}(s) = < A(s)^2 > \Omega \]

\[ \langle e^{isr} \rangle_\Omega = \frac{sinsr}{sr} \]

\[ < A(s) > = 4\pi \int_0^\infty \rho(r) \frac{\sin(sr)}{sr} r^2 dr = 4\pi \int_0^R \rho(r) \frac{\sin(sr)}{sr} r^2 dr = \]

\[ = \frac{4\pi}{s} \int_0^R \sin(sr) r dr = (\text{use partial integration}) = \]

\[ = \frac{4\pi}{s} \left( - \frac{R\cos(sr)}{s} + \left[ \frac{\sin(sr)}{s^2} \right]^R_0 \right) = \frac{4\pi}{s} \left( - \frac{R\cos(sr)}{s} + \frac{\sin(sR)}{s^2} \right) = \]

\[ = \frac{4\pi}{3} R^3 \frac{3(\sin(sR) - sR\cos(sR))}{(sR)^3} = V\Phi(sR) \]
Ellipsoid of revolution

\[ P(s) = \int_0^1 \Phi^2 [sR \left( 1 + x^2(\varepsilon^2 - 1) \right)]^{\frac{1}{2}} dx \]

P(q): Ellipsoid of revolution

prolate ellipsoid of revolution 1:1:3 (- - - - -)

oblate ellipsoid of revolution 1:1:0.2 (-----)

Prolates \((R,R,\varepsilon R)\)
\(\varepsilon > 1\)

Oblates \((R,R,\varepsilon R)\)
\(\varepsilon < 1\)
Bacteriophage T7 is a large bacterial virus with MM of 56 MDa consisting of an icosahedral protein capsid (diameter of about 600Å) that contains a double-stranded DNA molecule.

Instrumental smearing is routinely included in SANS data analysis.
Core-shell particles

\[ A(s) = [\Delta \rho_{\text{shell}} V_{\text{out}} \Phi(sR_{\text{out}}) - (\Delta \rho_{\text{shell}} - \Delta \rho_{\text{core}}) V_{\text{in}} \Phi(sR_{\text{in}})] \]

Where

\[ V_{\text{out}} = 4\pi R_{\text{out}}^{3/3} \quad \text{and} \quad V_{\text{in}} = 4\pi R_{\text{in}}^{3/3} \]

\[ \Phi(x) = \frac{3(sin(x) - xcos(x))}{x^3} \]

\( \Delta \rho_{\text{core}} \) – the excess scattering length density of the core

\( \Delta \rho_{\text{shell}} \) – the excess scattering length density of the shell
$P(s)=4 \int_0^1 \frac{J_1^2[sR(1-x^2)^{1/2}]}{[sR(1-x^2)^{1/2}]^2} S^2(sHx/2) \, dx$

$J_1(x)$ is the Bessel function of the first order and the first kind

$S(x)=\sin(x)/x$
Form factors of spheres and cylinders
Fitting data using geometrical bodies

Primus qt interface

Primus interface
Library of form-factors from geometrical bodies and polymer systems

SASFIT software
(J.Kohlbrecher, I.Bressler, PSI)


Jan Skov Pedersen
*Monte Carlo Simulation Techniques Applied in the Analysis of Small-Angle Scattering Data from Colloids and Polymer Systems* in *Neutrons, X-Rays and Light*
P. Lindner and Th. Zemb (Editors) 2002 Elsevier Science B.V.
p. 381

Jan Skov Pedersen
*Modelling of Small-Angle Scattering Data from Colloids and Polymer Systems* in *Neutrons, X-Rays and Light*
P. Lindner and Th. Zemb (Editors) 2002 Elsevier Science B.V.
p. 391
Scattering from mixtures (shape polydispersity)

The scattering is proportional to that of a single particle averaged over all orientations, which allows one to determine size, shape and internal structure of the particle at low (1-10 nm) resolution.

For equilibrium and non-equilibrium mixtures, solution scattering permits to determine the number of components and, given their scattering intensities $I_k(s)$, also the volume fractions.

$$I(s) = \sum_k v_k I_k(s)$$
Program OLIGOMER for SAXS analysis

Input parameters: 1) experimental data file (ASCII file *.dat)
2) form-factor file with the scattering from the components (can be easily prepared by FFMAKER)

\[ I(s) = \sum_k \nu_k I_k(s) \]

Output parameters: 1) the fit to experimental data (*.fit file)
2) the volume fractions of the components (in oligomer.log)

OLIGOMER can be launched in batch mode for multiple data sets:

oligomer.exe -ff formfactor.dat -dat hp*.dat -un 2 -smax 0.25

FFMAKER as pre-tool for OLIGOMER

To quickly create form-factor file from pdb files and/or from scattering data files (either from ASCII *.dat files or from GNOM output files where desmeared curve will be taken for intensity)

Batch mode:

ffmaker 1.dat 2.dat -undat 2 3.out -unout 2

ffmaker *.pdb m1.dat -smax 0.3 -ns 201 -lmmax 20

ffmaker 6lyz.pdb *.dat -sgrid m2.dat

Ab initio and rigid body analysis of the dimeric H(C) domain using the structure of the monomer in the crystal (1FV2) and accounting that the mutant Cys869Ala remains always monomeric yield a unique model of the dimer

Tricorn protease is a major component in the cleavage of oligopeptides produced by the proteasome.

Tricorn appeared to be a multifaceted system in solution.

The estimated molecular mass of the particles (380 kDa) was significantly lower than the theoretical value of 720 kDa tricorn hexamer, suggesting partial dissociation of the tricorn hexamers in solution.

SAXS data were fitted by a linear combination of the scattering from tricorn monomers (53%), dimers (14%) and hexamers (33%) using OLIGOMER.

Studies of adrenodoxin (Adx) : cytochrome c (Cc) complex by SAXS and NMR

Adx is involved in steroid hormone biosynthesis by acting as an electron shuttle between adrenodoxin reductase and cytochromes.

Solutions of native (WT) and cross-linked (CL) complex of Cc and Adx were measured by SAXS at different conditions:

a) solute concentration range from 2.4 to 24.0 mg/ml;

b) 10 mM Hepes / 20mM potassium phosphate (pH 7.4) buffer

c) with addition of NaCl (from 0 up to 300 mM).

Each protein has Molecular Mass (MM) of about 12.5 kDa.

For CL complex CcV28C and AdxL80C mutants were linked by a disulfide bond.

Studies of (Adx) : (Cc) complex formation

CL Complex

The experimental scattering from the CL complex does not depend on the solute concentration and addition of NaCl. It is compatible with 1:1 complex.

NMR structure of CL complex overlaps well with SAXS model.

The native complex strongly depends on the sample concentration and on the amount of NaCl in the buffer.

At high protein concentration it forms heterotetramer with 2:2 stoichiometry, whereas at high salt concentration it dissociates into two individual proteins.

Studies of (Adx) : (Cc) complex formation
Native Complex

<table>
<thead>
<tr>
<th></th>
<th>Native complex, no salt</th>
<th>CL complex</th>
</tr>
</thead>
<tbody>
<tr>
<td>c, mg/ml</td>
<td>24</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>2.4</td>
</tr>
<tr>
<td></td>
<td>3-12</td>
<td></td>
</tr>
<tr>
<td>R_g, Å</td>
<td>28.3±0.7</td>
<td>28.3±0.7</td>
</tr>
<tr>
<td></td>
<td>26.5±0.5</td>
<td>24.4±0.7</td>
</tr>
<tr>
<td></td>
<td>21.4±0.5</td>
<td></td>
</tr>
<tr>
<td>D_max, Å</td>
<td>90±5</td>
<td>90±5</td>
</tr>
<tr>
<td></td>
<td>90±5</td>
<td>80±5</td>
</tr>
<tr>
<td></td>
<td>80±5</td>
<td></td>
</tr>
<tr>
<td>V_p, 10^3 Å³</td>
<td>63±6</td>
<td>52±5</td>
</tr>
<tr>
<td></td>
<td>43±5</td>
<td>35±4</td>
</tr>
<tr>
<td></td>
<td>42±5</td>
<td></td>
</tr>
<tr>
<td>MM, kDa</td>
<td>44±5</td>
<td>42±5</td>
</tr>
<tr>
<td></td>
<td>35±4</td>
<td>25±4</td>
</tr>
<tr>
<td></td>
<td>22±3</td>
<td></td>
</tr>
<tr>
<td>V_mon,%</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>6±5</td>
<td>24±5</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>V_dim,%</td>
<td>0</td>
<td>8±5</td>
</tr>
<tr>
<td></td>
<td>25±5</td>
<td>24±5</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>V_tri,%</td>
<td>48±5</td>
<td>47±5</td>
</tr>
<tr>
<td></td>
<td>54±5</td>
<td>52±5</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>V_tet,%</td>
<td>52±5</td>
<td>45±5</td>
</tr>
<tr>
<td></td>
<td>15±5</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

OLIGOMER fits

Studies of (Adx) : (Cc) complex formation
Native Complex

Oligomerization behavior of the native complex in solution indicates a *stochastic nature* of complex formation. The native Adx/Cc is entirely dynamic and can be considered as a *pure encounter complex*.

**OLIGOMER fits**

The ensemble of native Adx:Cc complex structures from the PCS simulation.

More examples on polydisperse systems

Dynamic equilibria between monomers and higher oligomers (dissociation of multimers)

Dynamic equilibria between bound and free components for low-affinity transient complexes

The structures of the components are not known and/or the samples remain polydisperse at any conditions

**GASBORMX** (*ab initio* modelling) and **SASREFMX** (rigid body modelling) can take into account the polydispersity and restore the 3D models together with the volume fractions of the components

Ab initio modeling of partially dissociated multimers

Oligomer model

Asymmetric part (monomer)

Intensity

GASBORMX/SASREFMX: Oligomeric mixtures

Linear combination

Singular value decomposition (SVD)

For model-independent analysis of multiple scattering data sets from polydisperse systems, singular value decomposition (SVD) (Golub & Reinsch, 1970) can be applied.

The matrix \( A = \{A_{ik}\} = \{I(k)(s_i)\}, \)
(i = 1, \ldots, N, k = 1, \ldots, K,

where \( N \) is number of experimental points in the scattering curve and \( K \) is the number of data sets) is represented as \( A = U*S*V^T \), where the matrix \( S \) is diagonal, and the columns of the orthogonal matrices \( U \) and \( V \) are the eigenvectors of the matrices \( A*A^T \) and \( A^T*A \), respectively.
Singular value decomposition (SVD)

$$A = U * S * V^T$$

$$U * U^T = I$$

$$V * V^T = I$$

The matrix **U** yields a set of so-called left singular vectors, i.e. orthonormal basic curves **U(k)(s_i)**, that spans the range of matrix **A**, whereas the diagonal of **S** contains their associated singular values in descending order (the larger the singular value, the more significant the vector).
Singular value decomposition (SVD)

The number of significant singular vectors in SVD (i.e. 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).

SVD method has found wide-ranging applications:

* Spectrum analysis.
* Image processing and compression.
* Information Retrieval.
* Molecular dynamics.
* Analysis of gene expression data.
* Small-angle Scattering
* etc.
The program SVDPLOT computes the SVD from the active data sets in the PRIMUS toolbox and displays the singular vectors and singular values.

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 nonequilibrium mixtures [e.g. number of (un)folding or assembly intermediates].

\[
I_i(s) = \sum_{j=1}^{j=N} \lambda_{ij}(s)V_j(s)
\]

\[
\delta I_i(s) = I_i(s) - \sum_{j=1}^{j=p} \lambda_{ij}(s)V_j(s)
\]
Program SVDPLOT for SAXS analysis

PRIMUS: Number of independent components

Mixture of monomers and dimers
PRIMUS: Svdplot – singular value decomposition

For Q=0.02 (P=0.98), Ncomp = 2

Non-parametric test for the number of components.

In this test shapes of left-side singular vectors are tested for randomness of sequencing positive and negative elements. The number of components equals the number of singular vectors which shape is non-random (non-noisy).

In this table:

"No."

ordinal number of the singular vector. All the vectors are arranged correspondingly to singular values of the data matrix arranged in decreasing order.

"criterion":

testing statistics.

"Pr":

confidence level. In the corresponding column two numbers (columnwise pairs) are left and right critical values.

If the criterium lies outside critical values then the shape of the vector is NONrandom.

<table>
<thead>
<tr>
<th>No.</th>
<th>criterion</th>
<th>P=0.95</th>
<th>P=0.98</th>
<th>P=0.99</th>
<th>P=0.999</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-18.92</td>
<td>-2.017</td>
<td>-2.964</td>
<td>-2.633</td>
<td>-3.140</td>
</tr>
<tr>
<td>3</td>
<td>-1.185</td>
<td>-2.012</td>
<td>-2.379</td>
<td>-2.628</td>
<td>-3.142</td>
</tr>
<tr>
<td>4</td>
<td>3.928</td>
<td>-2.012</td>
<td>-2.378</td>
<td>-2.628</td>
<td>-3.143</td>
</tr>
</tbody>
</table>

For Q=0.02 (P=0.98), Ncomp = 2

Press to continue

Mixture of monomers and dimers
hNGF oligomeric state studied by SAXS

Mixture of dimers (D) and dimers of dimers (DD)


*Biophys. J* 108, 687-697
Scattering from **mixtures**
(size polydispersity)

\[ I(s) = \int_{R_{\text{min}}}^{R_{\text{max}}} V^2(R)N(R)i_0(sR)dR \]

Main structural task is determination of the size distribution function \( N(R) \) for a given form factor \( i_0(x) \)
Interparticle interactions
(concentration effects in protein solutions)

Ideal solution of particles
(diluted solutions)

Repulsive particle interactions

Attractive particle interactions
Interparticle interactions

For spherically symmetrical particles

\[ I(c, s) \approx I(0, s) \ast S(c, s) \]

form factor of the particle  
structure factor of the solution

Still valid for globular particles though over a restricted s-range

\( S(c,s) \) is related to the probability distribution function of inter-particle distances, i.e. *pair correlation* function \( g(r) \)
Interparticle interactions
(experimental structure factor)

\[ S(c, s) = \frac{c_0 I_{\text{exp}}(c, s)}{c I(c_0, s)} \]

The structure factor can be obtained from the ratio of the experimental intensity at a concentration \( c \) to that obtained by extrapolation to infinite dilution or measured at a sufficiently low concentration \( c_0 \) where all correlations between particles have vanished.
Interparticle interactions
High concentration studies of IgC2 antibody

The interactions between molecules depend on the buffer composition. The addition of NaCl changes attractive interactions (observed in normal buffer) to repulsive ones.

Computation of structure factor from interaction potentials

Excluded volume ‘repulsive’ interactions (‘hard-sphere’)

Short range attractive van der Waals interaction (‘stickiness’)

Electrostatic repulsive interaction (effective Debye-Hückel potential)

\[ v_e(r) = \frac{Q^2}{\varepsilon} \frac{e^{-\kappa(r-\sigma)}}{r} \]
SAXS/SANS studies on concentrated lysozyme solutions

Stradner et al. (2004) Nature reported that the position of the low-angle interference peak in small-angle x-ray and neutron scattering (SAXS and SANS) patterns from lysozyme solutions was essentially independent of the protein concentration and attributed these unexpected results to the presence of equilibrium clusters.

These experiments were repeated following the protein preparation protocols of Stradner et al. using several batches of lysozyme and exploring a broad range of concentrations, temperature and other conditions.

SAXS (EMBL X33 beamline)
SAXS (ESRF, ID02 beamline)
SANS (ILL, D22 beamline)

SAXS/SANS studies on concentrated lysozyme solutions

The new measurements revealed that the interference peak due to the repulsive interactions displayed a clear trend toward higher $q$ values with increasing protein concentration.

Several experimental sessions were performed in H2O and D2O buffers using different protein batches, different high resolution instruments and under varying experimental conditions (temperature, concentration, ionic strength, pH).

In all cases, the appearance and behavior of the interference peak is adequately and consistently described by the form and structure factors of individual lysozyme particles using an interaction potential involving short-range attraction and long-range repulsion.

Complex mixtures (size and shape polydispersity, interactions)

$$I(s) = \text{const} \sum_{k=1}^{K} \phi_k I_{k0}(s, R_{0k}, \Delta R_k) S_k(s, R_{ksh}, \eta_k, \tau_k)$$

Main structural task is determination of the volume fractions, average sizes, polydispersities and interactions by simulations or by non-linear fitting.
Originally written to analyse a morphological droplet-cylinder transition in AOT water-in-oil microemulsions to fit more than 500 scattering patterns at different physical and chemical conditions [1]

Now generalized to provide a restrained non-linear fit to the experimental data from polydisperse interacting mixtures of spheres, cylinders, dumbbells and ellipsoids

Scattering patterns from AOT microemulsions

- At low temperatures: mostly spherical particles
- At high temperatures: mostly long aggregates
- Without water: small reverse micelles
Temperature dependence, wo=25

Red: spherical droplets  Green: cylinders  Yellow: reverse micelles

Volume fractions

R₀, nm

<table>
<thead>
<tr>
<th>c</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
</tr>
</thead>
<tbody>
<tr>
<td>2%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

t, °C

<table>
<thead>
<tr>
<th>c</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
</tr>
</thead>
<tbody>
<tr>
<td>2%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
PRIMUS: non-linear analysis with MIXTURE

Volume fractions:
- spheres: 0.58
- AOT micelles: 0.17
- cylinders: 0.25
SAXS and EM study of Lymazine synthase

This enzyme catalyzes the formation of 6,7-dimethyl-8-ribityllumazine in the penultimate step of riboflavin biosynthesis.

The enzyme forms icosahedral capsids with a total molecular weight of about 960 kDa.

SAXS measurements were made for native and mutant enzyme species in different solvents and at different pH.

The formation of multiple assembly states was observed. They are interconvertible via equilibrium which is sensitive to solvent type and pH.

X.Zhang, P.Konarev, M.Petoukhov, D.Svergun et.al. JMB (2006) 362, 753-770
SAXS data from Lumazine synthase

SVD analysis yielded that the equilibrium mixtures for LSBS and LSAQ data contain five major components.
Lymazine synthase data analysis

WT, phosphate buffer

WT, Tris buffer

pH 7

pH 10

WT, Borate buffer

MIXTURE fits

Lymazine synthase data analysis

The system was successfully described by 5 components: complete and incomplete small capsids ($T=1$) complete and incomplete big capsids ($T=3,4$) free facets.

The data show that multiple assembly forms are a general feature of lumazine synthases.

X.Zhang, P.Konarev, M.Petoukhov, D.Svergun et.al. JMB (2006) 362, 753-770
Restoring the shape of unknown component from heterogeneous mixtures

Available algorithms

SAS data decomposition using MCR-ALS approach, program COSMiCS


Evolving Factor Analysis (EFA) for SEC-SAXS data.


Let us have N scattering curves collected from an evolving system (e.g. time series). In the beginning, we have the state with known intensity (e.g. monomer); at the end, we have also defined state (e.g. big aggregate). Quite often the situation is that there is an intermediate, whose scattering curve and structure is unknown.
The problem formulation for evolving system

- The scattering intensity at any (k-th) time point is a linear combination

\[ I_k(s) = v_{mk}I_m(s) + v_{ak}I_a(s) + v_{ik}I_i(s), \quad v_{mk} + v_{ak} + v_{ik} = 1 \]

The idea is to construct a shape that yields the intensity \( I_i(s) \) providing the best global fit (the overall \( \chi^2 \) over all observed data from the mixtures).
**Algorithm implementation (DAMMIX)**

- Dammif interface was modified to take multiple data sets

- The functional (R-factor) to be minimized was changed to compute the R-factor ($\chi^2$) over multiple curves

- At each SA step the volume fractions of the components are evaluated using the non-negative linear least-squares method (like in Oligomer)

- The functional is composed from the overall $\chi^2$ value plus the penalties for compactness/looseness and the penalty for the minimum threshold of the volume fraction for the intermediate
Restoring the shape of unknown component from heterogeneous mixtures

Possible practical cases (SAS data):

Evolving systems with unknown intermediate state (kinetic time series, studies of fibril formation, etc.)

Diluted oligomeric mixtures with unknown component studied at different conditions (pH, temperature, protein concentration, buffer composition, addition of ligand, etc.)

Multiple assembly states (virus-like structures, icosahedral capsids, formation of nanoparticles)

Program DAMMIX – combination of DAMMIN and OLIGOMER algorithms

P.V.Konarev & D.I.Svergun (2018) IUCr J., 5, 123
Fibrillation of insulin

5 g/l 20% acetic acid 0.5M NaCl 45°C

Growth rate of fibrils is proportional to volume fraction of intermediates

DAMMIX examples (insulin amyloid fibrils)

DAMMIX model (in green)
Experimental model from the paper (in magenta)

Restored volume fractions
DAMMIX examples
(Nerve growth factor NGF)

P.V. Konarev & D.I. Svergun (2018) IUCr J., 5, 123
DAMMIX examples (Lumazine synthase)

DAMMIX was able to find the presence of free facets within the mixture of big and small capsids.

P.V.Konarev & D.I.Svergun (2018) IUCr J., 5, 123
POLYSAS: interactive graphical program for analysis of polydisperse systems and multiple data sets

Overall parameters for multiple data sets
Size distributions for polydisperse systems
Volume fractions of components in equilibrium mixtures
Interparticles interactions in polydisperse systems

POLYSAS: interactive analysis

Ataxin-1 is a human protein responsible for ataxia type 1, a hereditary disease associated with protein aggregation and misfolding. The AXH domain of Ataxin-1 forms a globular dimer in solution and displays a dimer of dimers arrangement in the crystal asymmetric unit. In solution, the domain is present as a complex equilibrium mixture of monomeric, dimeric, and higher molecular weight species. This behavior, together with the tendency of the AXH fold to be trapped in local conformations, and the multiplicity of protomer interfaces, makes the AXH domain an unusual example of a chameleon protein.

Complex equilibrium mixture of ataxin-1 in solution

de Chiara, C., Rees, M., Menon, R.P., et.al. (2013) *Biophys J.* 104, 1304
Conclusions

- ATSAS package allows one to quantitatively analyze interacting and polydisperse systems and mixtures:
  - to determine volume fractions of oligomers (OLIGOMER)
  - to account for polydispersity in 3D modelling algorithms (GASBORMX, SASREFMX)
  - to make model-independent estimation of significant components for systems measured at different conditions or for kynetic processes (SVDPLOT)
  - to quantitatively characterize systems with size and shape polydispersity as well as systems with interparticle interactions (MIXTURE)
  - to restore the shapes of intermediates in evolving systems (DAMMIX)
  - to interactively process multiple data sets from polydisperse systems (POLYSAS)
  - to estimate conformational ensembles of flexible systems (EOM)
Thank you!